**National Junior College**
**2016 – 2017 H2 Further Mathematics**
**Topic F7: Matrices and Linear Spaces (Lecture Notes)**

*Key Questions to Answer:*

1. How do we use matrices to represent a set off linear equations?
2. What are the common operations on matrices?
3. How do we find the determinant of a $2\times 2$ or $3\times 3$ matrix?
4. How do we find the inverse of a non-singular $2\times 2$ or $3\times 3$ matrix?
5. How do we use matrices to solve a set of linear equations? What is the geometrical interpretation of the solution?
6. What is a linear space? What is a subspace?
7. What are the axioms for a linear space?
8. What is a span? What is linear independence?
9. How do we find the basis and dimension of a linear space?
10. How do we find the column space, row space, range space and null space of a matrix?
11. What is the rank of a square matrix? What is the relation between the rank, dimension of null space and the order of the matrix?
12. What are linear transformations?
13. What are the eigenvalues and eigenvectors of a $2\times 2$ or $3\times 3$ matrix?
14. How do we diagonalize a square matrix? What are the applications of diagonalization?

## §1 System of Linear Equations

Systems of linear equations arise in a wide variety of applications, such as polynomial curve fitting, network analysis and optimisation. You may refer to **Appendix III** for more details.

### 1.1 Linear Systems

**Definition**

A ***system of m linear equations in n unknown $x_1, x_2, x_3, \ldots, x_n$*** is a set of $m$ linear equations each in $n$ unknowns:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m, \end{aligned} \qquad (*)$$

where $a_{ij}$ and $b_i$, $1 \le i \le m$, $1 \le j \le n$ are constants.

A sequence of numbers $s_1, s_2, \ldots, s_n$ (or $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$) is called ***a solution*** of the system (*) if every equation in the system is satisfied when we substitute $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$.

**Example 1.1.1**

Verify that $x = 1$, $y = 2$ and $z = -2$ is *a solution* of the linear system

$$\begin{aligned} x + y - z &= 5 \\ x - 3z &= 7. \end{aligned}$$

Determine whether $x = 2$, $y = 3$ and $z = 0$ is also a solution of the system. Suggest another solution of the system.

*Solution:*

Substitute $x = 1$, $y = 2$ and $z = -2$ into both equations, since $1 + 2 - (-2) = 5$ and $1 - 3(-2) = 7$, it is a solution of the linear system.

Since $2 - 3(0) = 2 \ne 7$, $x = 2$, $y = 3$ and $z = 0$ is not a solution of the system.

Another solution can be $x = 4$, $y = 0$ and $z = -1$ (not unique)

**Example 1.1.2**

Suggest the number of solution(s) of each of the following linear systems:

**(a)** $x+y=2$, $x-y=4$ **(b)** $x-y=1$, $-2x+2y=6$ **(c)** $x+2y=1$, $2x+4y=2$

*Solution:*

**(a)** Exactly one solution **(b)** No solution **(c)** Infinitely many solutions

**Theorem 1.1.1**

Every system of linear equations has either no solution, exactly one solution or infinitely many solutions. (There are no other possibilities)

- The theorem is not true if the equations are not all linear. Can you give an example?
- For a system of linear equations in 2 unknowns, what is the geometrical interpretation of the theorem?
- For a system of linear equations in 3 unknowns, what is the geometrical interpretation of the theorem?

**Definition**

If a system of equations has no solution, they we say that it is ***inconsistent***; if the system has at least one solution, they we say that it is ***consistent***.

In **Example 1.1.2**, **(a)** and **(c)** are consistent, but **(b)** is consistent.

**Example 1.1.3**

Solve the following linear system by *elimination*

$$\begin{aligned} 3x-2y&=1 \\ x+4y&=6 \end{aligned}$$

*Solution:*

$$\begin{aligned} x+4y&=6 \\ 3x-2y&=1 \end{aligned} \qquad (1)$$

$$\begin{aligned} x+4y&=6 \\ -14y&=-17 \end{aligned} \qquad (2)$$

$$\begin{aligned} x+4y&=6 \\ y&=\frac{17}{14} \end{aligned} \qquad (3)$$

By backward substitution, we obtain the solution of the linear system: $x=\frac{8}{7}$ and $y=\frac{17}{14}$.

**Example 1.1.4**

Solve the following linear system by *elimination*

$$\begin{array}{llll} x & -3y & & =2 \\ -x & +y & +5z & =2 \\ 2x & -5y & +z & =0 \end{array}$$

*Solution:*

$$\begin{array}{llll} x & -3y & & =2 \\ & -2y & +5z & =4 \\ 2x & -5y & +z & =0 \end{array} \qquad (1)$$

$$\begin{array}{llll} x & -3y & & =2 \\ & -2y & +5z & =4 \\ & y & +z & =-4 \end{array} \qquad (2)$$

$$\begin{array}{llll} x & -3y & & =2 \\ & y & +z & =-4 \\ & -2y & +5z & =4 \end{array} \qquad (3)$$

$$\begin{array}{llll} x & -3y & & =2 \\ & y & +z & =-4 \\ & & 7z & =-4 \end{array} \qquad (4)$$

$$\begin{array}{llll} x & -3y & & =2 \\ & y & +z & =-4 \\ & & z & =-\frac{4}{7} \end{array} \qquad (5)$$

By backward substitution, we obtain the solution of the linear system:

$$x=-\frac{58}{7},\ y=-\frac{24}{7} \text{ and } z=-\frac{4}{7}.$$

- In the processes of solving **Example 1.1.3** and **Example 1.1.4**, what types of operations have we performed in each step?

Note that the method of elimination is to simplify a system of linear equations to another system of linear equations that has <u>exactly the same set of solution(s)</u>, but is easier to solve.

In the method of elimination, we perform the following <u>three</u> types of operations:

1. Multiply an equation through by a <u>nonzero</u> constant.
2. Interchange two equations.
3. Add a multiple of one equation to another.

## 1.2 Gaussian and Gauss-Jordan Elimination

**Definition**

Given a linear system (*) above, the rectangle array of numbers

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$

is called the ***augmented matrix*** of the linear system (*).

**Example 1.2.1**

Write down the augmented matrix of each of the following linear systems:

**(a)**
$$\begin{array}{llll} -2x & & +z & =5 \\ 2x & +3y & -4z & =7 \\ 3x & +2y & & =3 \end{array}$$

**(b)**
$$\begin{array}{l} x=1 \\ y=2 \\ z=3 \end{array}$$

*Solution:*

**(a)** $\begin{pmatrix} -2 & 0 & 1 & 5 \\ 2 & 3 & -4 & 7 \\ 3 & 2 & 0 & 3 \end{pmatrix}$

**(b)** $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{pmatrix}$

**Definition**

Corresponding to the three types of operations in the method of elimination, the following operations on the rows of the augmented matrix are called ***elementary row operations***:

1. Multiply a row through by a nonzero constant.
2. Interchange two rows.
3. Add a multiple of one row to another row.

**Example 1.2.2**

Solve the linear system in **Example 1.1.4** by performing elementary row operations:

$$\begin{array}{llll} x & -3y & & =2 \\ -x & +y & +5z & =2 \\ 2x & -5y & +z & =0 \end{array}$$

*Solution:*

The augmented matrix of the linear system is $\begin{pmatrix} 1 & -3 & 0 & 2 \\ -1 & 1 & 5 & 2 \\ 2 & -5 & 1 & 0 \end{pmatrix}$.

$$\begin{pmatrix} 1 & -3 & 0 & 2 \\ -1 & 1 & 5 & 2 \\ 2 & -5 & 1 & 0 \end{pmatrix} \xrightarrow{R2+R1} \begin{pmatrix} 1 & -3 & 0 & 2 \\ 0 & -2 & 5 & 4 \\ 2 & -5 & 1 & 0 \end{pmatrix} \xrightarrow{R3+(-2)R1} \begin{pmatrix} 1 & -3 & 0 & 2 \\ 0 & -2 & 5 & 4 \\ 0 & 1 & 1 & -4 \end{pmatrix} \xrightarrow{R2\leftrightarrow R3} \begin{pmatrix} 1 & -3 & 0 & 2 \\ 0 & 1 & 1 & -4 \\ 0 & -2 & 5 & 4 \end{pmatrix}$$

$$\xrightarrow{R3+2R2} \begin{pmatrix} 1 & -3 & 0 & 2 \\ 0 & 1 & 1 & -4 \\ 0 & 0 & 7 & -4 \end{pmatrix} \xrightarrow{\frac{R3}{7}} \begin{pmatrix} 1 & -3 & 0 & 2 \\ 0 & 1 & 1 & -4 \\ 0 & 0 & 1 & -\frac{4}{7} \end{pmatrix}$$

By backward substitution, we obtain the solution of the linear system:

$$x = -\frac{58}{7},\ y = -\frac{24}{7} \text{ and } z = -\frac{4}{7}.$$

Consider the following two linear systems:

$$\begin{array}{ccccl} x & +2y & -z & +5w & = -1 \\ & y & +3z & -w & = 2 \\ & & z & +2w & = 3 \\ & & & w & = 1 \end{array} \qquad (1) \qquad\qquad \begin{array}{ccccl} x & & & & = 3 \\ & y & & & = 1 \\ & & z & & = 2 \\ & & & w & = 5 \end{array} \qquad (2)$$

The solution to (1) can be obtained by backward substitution, while the solution to (2) is immediate.

In solving a linear system by the method of elimination, the aim is to reduce the linear system (by performing the three operations stated in **Section 1.1**) to an equivalent system (having the same set of solution(s) as the original system) similar to (1), or to further reduce it to a system similar to (2).

The augmented matrices of the linear systems (1) and (2) are respectively

$$\begin{pmatrix} 1 & 2 & -1 & 5 & -1 \\ 0 & 1 & 3 & -1 & 2 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 5 \end{pmatrix}.$$

The first matrix is an example of a matrix in *row-echelon form*, while the second matrix is an example of a matrix in *reduced row-echelon form*.

**Definition**

A matrix is said to be in ***row-echelon form*** if it satisfies all the following properties:

1. If there are any rows that consist entirely of zeros, then they are grouped together at the bottom of the matrix.
2. If a row does not consist of entirely of zeros, then the first nonzero number in the row is a 1. We call this a ***leading 1***.
3. In any two successive rows that do not consists entirely of zeros, the leading 1 in the lower row occurs further to the right than the leading 1 in the higher row.

The matrix is said to be in ***reduced row-echelon form*** if, <u>in addition</u> to the above three properties, the following property is satisfied:

4. Each column that contains a leading 1 has zeros everywhere else in that column.

Here are some examples:

$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ is not in row-echelon form; $\begin{pmatrix} 0 & 1 & 2 & 6 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$ is in row-echelon form but not in reduced row-echelon form; $\begin{pmatrix} 1 & 0 & 7 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is in reduced row-echelon form.

- Does a given matrix **A** have a unique row-echelon form?
- Does a given matrix **A** have a reduced unique row-echelon form?

**Example 1.2.3**

Determine if each of the following matrices is in row-echelon form. For those matrices in row-echelon form, which are in reduced row-echelon form?

**(a)** $\begin{pmatrix} 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -3 \end{pmatrix}$ **(b)** $\begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ **(c)** $\begin{pmatrix} 1 & 0 & 3 & 4 \\ 0 & 1 & -2 & 5 \\ 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 0 \end{pmatrix}$ **(d)** $\begin{pmatrix} 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

*Solution:*

**(a)** Not in row-echelon form
**(b)** In row-echelon form but not in reduced row-echelon form
**(c)** Not in row-echelon form
**(d)** In reduced row-echelon form

**Example 1.2.4 (Linear system with a unique solution)**

The augmented matrix of a linear system in $(x, y, z)$ has been reduced to the given row-echelon form:

$$\begin{pmatrix} 1 & 2 & -1 & 2 \\ 0 & 1 & 3 & -1 \\ 0 & 0 & 1 & 4 \end{pmatrix}.$$

Solve the linear system.

*Solution:*

The corresponding linear system is

$$\begin{aligned} x \quad +2y \quad -z \quad &= 2 \\ y \quad +3z \quad &= -1 \\ z \quad &= 4 \end{aligned}$$

By backward substitution, we obtain the solution $x = 32$, $y = -13$ and $z = 4$.

**Example 1.2.5 (Linear system with infinitely many solutions)**

Write down all the solutions of $x + 2y - z = 3$.

*Solution*

Let $y = s$ and $z = t$, then $x = 3 - 2s + t$. Thus, all the solutions are $x = 3 - 2s + t$, $y = s$ and $z = t$, where $s, t \in \mathbb{R}$.

Note that the quantities $s$ and $t$ are called ***parameters***, and the set of all solutions expressed in terms of the parameters is called the ***general solution*** of the linear system.

**Example 1.2.6**

The augmented matrix of a linear system in $(x, y, z, w)$ has been reduced to the reduced-row echelon form:

$$\begin{pmatrix} 1 & 0 & 0 & 2 & -7 \\ 0 & 1 & 0 & 1 & 5 \\ 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Solve the linear system.

*Solution:*

The corresponding linear system is

$$\begin{aligned} x \quad\quad\quad\quad &+2w = -7 \\ y \quad\quad &+w = 5 \\ z &+3w = 1 \end{aligned}$$

The variables (unknowns) that corresponding to the leading 1's, namely $x$, $y$ and $z$, are called ***leading variables***. The non-leading variables ($w$ in this case) are called ***free variables***.

Solving for leading variables in terms of variables, we can assign any arbitrary value to the free variable $w$, say $t$, which then determines the values of the leading variable. Thus this linear system has infinitely many solutions given by

$$x = -7 - 2t,\ y = 5 - t,\ z = 1 - 3t,\ w = t, \text{ where } t \in \mathbb{R}.$$

**Definition**

The method of solving a linear system by reducing the corresponding augmented matrix to row-echelon form (respectively reduced row-echelon form) is unknown as ***Gaussian elimination*** (respectively ***Gauss-Jordan elimination***).

**Example 1.2.7**

Without using a calculator, solve the linear system

$$\begin{aligned} 3x + 4y - 2z + 13w &= 9 \\ x + 2y - 2z + 7w &= 5 \\ 2x + y + 4z + 6w &= -3 \end{aligned}$$

*Solution:*

We write down the augmented matrix of the linear system and then perform elementary row operations to reduce it to row-echelon form or reduced row-echelon form.

$$\begin{pmatrix} 3 & 4 & -2 & 13 & 9 \\ 1 & 2 & -2 & 7 & 5 \\ 2 & 1 & 4 & 6 & -3 \end{pmatrix} \xrightarrow{R1 \leftrightarrow R2} \begin{pmatrix} 1 & 2 & -2 & 7 & 5 \\ 3 & 4 & -2 & 13 & 9 \\ 2 & 1 & 4 & 6 & -3 \end{pmatrix} \xrightarrow[R3-R1\times 2]{R2-R1\times 3} \begin{pmatrix} 1 & 2 & -2 & 7 & 5 \\ 0 & -2 & 4 & -8 & -6 \\ 0 & -3 & 8 & -8 & -13 \end{pmatrix}$$

$$\xrightarrow{R2\times\left(-\frac{1}{2}\right)} \begin{pmatrix} 1 & 2 & -2 & 7 & 5 \\ 0 & 1 & -2 & 4 & 3 \\ 0 & -3 & 8 & -8 & -13 \end{pmatrix} \xrightarrow{R3+R2\times 3} \begin{pmatrix} 1 & 2 & -2 & 7 & 5 \\ 0 & 1 & -2 & 4 & 3 \\ 0 & 0 & 2 & 4 & -4 \end{pmatrix} \xrightarrow{R3\times\frac{1}{2}} \begin{pmatrix} 1 & 2 & -2 & 7 & 5 \\ 0 & 1 & -2 & 4 & 3 \\ 0 & 0 & 1 & 2 & -2 \end{pmatrix}.$$

The linear system corresponding to the row-echelon form is

$$\begin{aligned} x + 2y - 2z + 7w &= 5 \\ y - 2z + 4w &= 3 \\ z + 2w &= -2 \end{aligned}$$

which has the same set of solutions as the given linear system. Now $x, y$ and $z$ are the leading variables, and $w$ is the free variable. Let $w=t$, where $t$ is an arbitrary real number. By backward substitution, $z=-2-2t$, $y=3+2z-4w=-1-8t$, $x=5-2y+2z-7w=3+5t$.

Thus the general solution of the given linear system is

$$x=3+5t,\ y=-1-8t,\ z=-2-2t,\ w=t, \text{ where } t\in\mathbb{R}.$$

Alternatively, we can further reduce the row-echelon form to reduced row-echelon form:

$$\begin{pmatrix} 1 & 2 & -2 & 7 & 5 \\ 0 & 1 & -2 & 4 & 3 \\ 0 & 0 & 1 & 2 & -2 \end{pmatrix} \xrightarrow[R2+R3\times 2]{R1+R3\times 2} \begin{pmatrix} 1 & 2 & 0 & 11 & 1 \\ 0 & 1 & 0 & 8 & -1 \\ 0 & 0 & 1 & 2 & -2 \end{pmatrix} \xrightarrow{R1+R2\times(-2)} \begin{pmatrix} 1 & 0 & 0 & -5 & 3 \\ 0 & 1 & 0 & 8 & -1 \\ 0 & 0 & 1 & 2 & -2 \end{pmatrix}.$$

The corresponding linear system is now

$$\begin{array}{cccrl} x & & & -5w & =3 \\ & y & & +8w & =-1 \\ & & z & +2w & =-2 \end{array}$$

We will be able to obtain the same general solution by assigning $w=t$.

**Example 1.2.8 (Geometrical interpretation)**

The general solution of the system of linear equations

$$\begin{array}{cccl} x & +y & & =-1 \\ 2x & +y & +z & =3 \\ x & & +z & =4 \end{array}$$

is given by $x=4-t$, $y=-5+t$, $z=t$. What is the geometrical interpretation of the solution?

*Solution:*

The three planes $x+y=-1$, $2x+y+z=3$ and $x+z=4$ intersect in a common line, with vector equation $r=\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 4-t \\ -5+t \\ t \end{pmatrix}=\begin{pmatrix} 4 \\ -5 \\ 0 \end{pmatrix}+t\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$, $t\in\mathbb{R}$.

- What are the geometrical interpretations of the solutions of **Example 1.1.2** and **Example 1.1.4**?

### 1.3 Homogenous Linear Systems

**Definition**

A linear system of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= 0 \\ a_{21}x_2 + a_{22}x_2 + \ldots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= 0 \end{aligned}$$

is called a ***homogeneous linear system***.

Every homogeneous linear system is consistent, since $x_1 = 0$, $x_2 = 0$, …, $x_n = 0$ is a solution. This solution is called the ***trivial solution***; if there are <u>other</u> solutions, then they are called ***nontrivial solutions*** (i.e. a solution $x_1 = s_1$, $x_2 = s_2$, …, $x_n = s_n$ is a nontrivial solution if <u>at least one</u> of $s_1$, $s_2$, …, $s_n$ is not equal to 0).

**Example 1.3.1**

Find the solutions of the homogeneous systems

**(a)** $$\begin{aligned} x + 2y &= 0 \\ -x + 3y &= 0 \end{aligned}$$

**(b)** $$\begin{array}{lllll} x & +y & +z & +w & = 0 \\ x & & & +w & = 0 \\ x & +2y & +z & & = 0 \end{array}.$$

*Solution:*

**(a)** This homogeneous system has only one solution, which is the trivial solution $x = 0$, $y = 0$.

**(b)** Using Gauss-Jordan elimination, we obtain an equivalent linear system

$$\begin{array}{lllll} x & & & +w & = 0 \\ & y & & -w & = 0 \\ & & z & +w & = 0 \end{array}$$

Let $w = t$, where $t$ is an arbitrary real number. Then the general solution of the homogeneous linear system is $x = -t$, $y = t$, $z = -t$, $w = t$.

In **Example 1.3.1(a)**, the homogeneous system has only one solution (the trivial solution); whereas in **Example 1.3.1(b)**, the homogeneous system has infinitely many solutions.

**Theorem 1.3.1**

Every homogeneous system of linear equations with more unknowns than equations has infinity many solutions.

- In the context of a homogeneous system of one/two linear equations in three unknowns, how can we justify this theorem geometrically?

**Example 1.3.2**

Determine whether the homogeneous linear system has nontrivial solution

$$\begin{array}{cccl} x & +y & +3z & =0 \\ -x & +2y & +6z & =0 \quad \ldots(1) \\ 2x & -y & -3z & =0 \end{array}$$

*Solution:*

Perform elementary row operations on the augmented matrix:

$$\begin{pmatrix} 1 & 1 & 3 & 0 \\ -1 & 2 & 6 & 0 \\ 2 & -1 & -3 & 0 \end{pmatrix} \xrightarrow[R3+(-2)R1]{R2+R1} \begin{pmatrix} 1 & 1 & 3 & 0 \\ 0 & 3 & 9 & 0 \\ 0 & -3 & -9 & 0 \end{pmatrix} \xrightarrow{R3+R2} \begin{pmatrix} 1 & 1 & 3 & 0 \\ 0 & 3 & 9 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The corresponding homogeneous system $\begin{array}{cccl} x & +y & +3z & =0 \\ & 3y & +9z & =0 \end{array}$ has 3 unknowns and 2 equations. Hence the homogenous linear system has nontrivial solution by **Theorem 1.3.1**. Since it is equivalent to the homogeneous system (1), (1) also has nontrivial solution.

## §2 Matrices and Matrix Operations

### 2.1 Notation and Terminology

**Definition**

***A matrix*** is a rectangle array of numbers. We say that a matrix is of size $m$ by $n$ (written $m \times n$) if it has $m$ rows (the horizontal lines) and $n$ columns (the vertical lines).

A matrix with only one row is called a ***row matrix***, and a matrix with only one column is called a ***column matrix***.

The numbers in the array are called the ***entries*** in the matrix. The entry in the $i$th row and $j$th column of a matrix is called the $(i, j)$ ***entry*** of the matrix. A general $m \times n$ matrix is written as

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Note that $a_{ij}$ is the $(i, j)$ entry of the matrix **A**, commonly denoted by $(\mathbf{A})_{ij}$.

**Definition**

A matrix with $n$ rows and $n$ columns (so the number of rows = number of columns) is called a ***square matrix of order n***, the entries $a_{11}, a_{12}, ..., a_{nn}$ in the matrix below are said to be the ***main diagonal*** of **A**.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

The ***trace*** if a square matrix **A**, denoted by $\text{tr}(\mathbf{A})$, is defined to be the sum of all entries on the main diagonal of **A**.

For example, let

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 3 \\ 4 & 5 & -2 \\ 3 & 6 & 7 \end{pmatrix}.$$

Then the trace of the square matrix $A$ is $1+5+7=13$.

## 2.2 Operations on Matrices

**Definition**

If **A** and **B** are matrices of the same size, then the ***sum*** $\mathbf{A}+\mathbf{B}$ is the matrix obtained by adding the entries of **B** to the corresponding entries of **A**; and the ***difference*** $\mathbf{A}-\mathbf{B}$ is the matrix obtained by subtracting the entries of **B** from the corresponding entries of **A**.

In matrix notation,

$$\left(\mathbf{A}\pm\mathbf{B}\right)_{ij}=\left(\mathbf{A}\right)_{ij}\pm\left(\mathbf{B}\right)_{ij}.$$

**Definition**

If **A** is any matrix and $k$ is any scalar (real number), then the ***scalar*** multiple of **A**, by $k$, denoted by $k\mathbf{A}$, is the matrix obtained by multiplying each entry of **A** by $k$.

In matrix notation,

$$\left(k\mathbf{A}\right)_{ij}=k\left(\mathbf{A}\right)_{ij}.$$

For example, let $\mathbf{A}=\begin{pmatrix}1 & -2 & 4\\ 2 & -1 & 3\end{pmatrix}$ and $\mathbf{B}=\begin{pmatrix}0 & 2 & -4\\ 1 & 3 & 1\end{pmatrix}$.

Then $\mathbf{A}+\mathbf{B}=\begin{pmatrix}1 & 0 & 0\\ 3 & 2 & 4\end{pmatrix}$, $\mathbf{A}-\mathbf{B}=\begin{pmatrix}1 & -4 & 8\\ 1 & -4 & 2\end{pmatrix}$ and $\left(-2\right)\mathbf{A}=\begin{pmatrix}-2 & 4 & -8\\ -4 & 2 & -6\end{pmatrix}$.

**Definition**

If **A** is an $m\times r$ matrix and **B** is an $r\times n$ matrix, then the ***product*** **AB** is the $m\times n$ matrix whose entries are determined as follows:

$$\left(\mathbf{AB}\right)_{ij}=\left(\mathbf{A}\right)_{i1}\left(\mathbf{B}\right)_{1j}+\left(\mathbf{A}\right)_{i2}\left(\mathbf{B}\right)_{2j}+\ldots+\left(\mathbf{A}\right)_{ir}\left(\mathbf{B}\right)_{1r}.$$

- For the product **AB** to be defined, the number of columns of **A** must be equal to the number of rows of **B**.

**Example 2.2.1**

Let $\mathbf{A}=\begin{pmatrix}3 & 0\\ -1 & 2\\ 1 & 1\end{pmatrix}$ and $\mathbf{B}=\begin{pmatrix}1 & 4 & 2\\ 3 & 1 & 5\end{pmatrix}$. Compute **AB** and **BA**.

*Solution:*

$\mathbf{AB}=\begin{pmatrix}3 & 12 & 6\\ 5 & -2 & 8\\ 4 & 5 & 7\end{pmatrix}$ and $\mathbf{BA}=\begin{pmatrix}1 & 10\\ 13 & 7\end{pmatrix}$.

In **Example 2.2.1**, multiplying **A** with the first, second and third columns of **B**, we obtain respectively the first, second and third columns of **AB**, i.e.

$$\mathbf{A}\begin{pmatrix}1\\3\end{pmatrix}=\begin{pmatrix}3\\5\\4\end{pmatrix}\quad \mathbf{A}\begin{pmatrix}4\\1\end{pmatrix}=\begin{pmatrix}12\\-2\\5\end{pmatrix}\quad \mathbf{A}\begin{pmatrix}2\\5\end{pmatrix}=\begin{pmatrix}6\\8\\7\end{pmatrix}.$$

Similarly, multiplying the first, second, third rows of $A$ with the matrix **B**, we obtain respectively the first, second and third rows of **AB**, i.e.

$$\begin{pmatrix}3 & 0\end{pmatrix}\mathbf{B}=\begin{pmatrix}3 & 12 & 6\end{pmatrix}\quad \begin{pmatrix}-1 & 2\end{pmatrix}\mathbf{B}=\begin{pmatrix}5 & -2 & 8\end{pmatrix}\cancel{\begin{pmatrix}1 & 1\end{pmatrix}\mathbf{B}=\begin{pmatrix}4 & 5 & 7\end{pmatrix}}\quad \begin{pmatrix}1 & 1\end{pmatrix}\mathbf{B}=\begin{pmatrix}4 & 5 & 7\end{pmatrix}$$

In general, if **A** and **B** are matrices such that **AB** is defined, then

$j$ th column of **AB** = **A** ($j$ th column of **B**), and

$i$ th row of **AB** = ($i$ th row of $A$) **B**.

**Example 2.2.2**

Let $\mathbf{A}=\begin{pmatrix}1 & 3 & 2\\-1 & 0 & 4\end{pmatrix}$ and $\mathbf{B}=\begin{pmatrix}1 & 4 & 3\\0 & -2 & 5\\7 & 1 & -1\end{pmatrix}$, find the 2nd column of **AB**.

*Solution:*

2nd column of **AB** = $\begin{pmatrix}1 & 3 & 2\\-1 & 0 & 4\end{pmatrix}\begin{pmatrix}4\\-2\\1\end{pmatrix}=\begin{pmatrix}0\\0\end{pmatrix}$.

**Matrix Form of a Linear System**

Now a linear system

$$\begin{aligned}a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n&=b_1\\a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n&=b_2\\&\vdots\\a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n&=b_m,\end{aligned}\qquad (*)$$

can be rewritten in the form

$$\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\a_{21} & a_{22} & \cdots & a_{2n}\\\vdots & \vdots & & \vdots\\a_{m1} & a_{m2} & \cdots & a_{mn}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\begin{pmatrix}b_1\\b_2\\\vdots\\b_m\end{pmatrix}.$$

Thus the original system of $m$ equations in $n$ unknowns can be replaced by a single matrix equation $\mathbf{Ax}=\mathbf{b}$. The matrix **A** is called the ***coefficient matrix*** of the linear system.

Do not confuse the matrix form of the linear system (*) with its augmented matrix, which is $(\mathbf{A}\ \mathbf{b})$.

**Example 2.2.3**

Write down the matrix equation of the linear system

$$\begin{array}{rrrl} 2x & & +z & =5 \\ -x & +4y & -3z & =1 \end{array}.$$

*Solution:*

$$\begin{pmatrix} 2 & 0 & 1 \\ -1 & 4 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}=\begin{pmatrix} 5 \\ 1 \end{pmatrix}.$$

| **Definition** |
|---|
| If **A** is any $m\times n$ matrix, then the transpose of **A**, denoted by $\mathbf{A}^T$, is defined to be the $n\times m$ matrix that results from interchanging the rows and columns of **A**, i.e. $\left(\mathbf{A}^T\right)_{ij}=\left(\mathbf{A}\right)_{ji}$. |

For example, if $\mathbf{A}=\begin{pmatrix} 2 & 3 \\ 1 & 4 \\ -5 & 6 \end{pmatrix}$, then $\mathbf{A}^T=\begin{pmatrix} 2 & 1 & -5 \\ 3 & 4 & 6 \end{pmatrix}$.

### 2.3 Properties of Matrix Operations

For any real numbers $a$, $b$ and $c$, we know that

$a+b=b+a$ [commutative Law for Addition]

$a+\left(b+c\right)=\left(a+b\right)+c$ [associative Law for Addition]

| **Theorem 2.3.1** |
|---|
| Let **A**, **B** and **C** be $m\times n$ matrices, then<br>$\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$ [commutative law for addition]<br>$\mathbf{A}+\left(\mathbf{B}+\mathbf{C}\right)=\left(\mathbf{A}+\mathbf{B}\right)+\mathbf{C}$ [associative law for addition] |

Because of the associate law for matrix addition, we may write $\mathbf{A}+\mathbf{B}+\mathbf{C}$ without ambiguity if **A**, **B** and **C** have the same size. Similarly for the sum of more than 3 matrices.

With regard to matrix multiplication, some, <u>but not all</u>, properties of real number multiplication carry over to matrix multiplication:

| **Theorem 2.3.2** |
|---|
| Assume **A**, **B** and **C** are matrices of appropriate sizes so that the indicated operations are defined, then<br>$\mathbf{A}\left(\mathbf{BC}\right)=\left(\mathbf{AB}\right)\mathbf{C}$ [associative law for multiplication]<br>$\mathbf{A}\left(\mathbf{B}+\mathbf{C}\right)=\mathbf{AB}+\mathbf{AC}$ [left distributive law]<br>$\left(\mathbf{A}+\mathbf{B}\right)\mathbf{C}=\mathbf{AC}+\mathbf{BC}$ [right distributive law] |

*Proof for the associative law for multiplication:*

Let $\mathbf{A}=\left(a_{ij}\right)$, $\mathbf{B}=\left(b_{ij}\right)$ and $\mathbf{C}=\left(c_{ij}\right)$ be matrices of sizes $m\times n$, $n\times r$ and $r\times s$ respectively.

The matrix **BC** is of size $n\times s$, so the matrix **A**(**BC**) is of size $m\times s$. The matrix **AB** is of size $m\times r$, so the matrix (**AB**)**C** is of size $m\times s$. Hence **A**(**BC**) and (**AB**)**C** have the same size

Next we show that the any corresponding entries of the two matrices are equal:

$$
\begin{aligned}
\left[\mathbf{A}\left(\mathbf{BC}\right)\right]_{ij} &= a_{i1}\left(\mathbf{BC}\right)_{1j}+a_{i2}\left(\mathbf{BC}\right)_{2j}+\ldots+a_{in}\left(\mathbf{BC}\right)_{nj}\\
&= a_{i1}\left(b_{11}c_{1j}+b_{12}c_{2j}+\ldots+b_{1r}c_{rj}\right)\\
&\quad +a_{i2}\left(b_{21}c_{1j}+b_{22}c_{2j}+\ldots+b_{2r}c_{rj}\right)\\
&\quad +\ldots\\
&\quad +a_{in}\left(b_{n1}c_{1j}+b_{n2}c_{2j}+\ldots+b_{nr}c_{rj}\right)\\
&= a_{i1}b_{11}c_{1j}+a_{i1}b_{12}c_{2j}+\ldots+a_{i1}b_{1r}c_{rj}\\
&\quad +a_{i2}b_{21}c_{1j}+a_{i2}b_{22}c_{2j}+\ldots+a_{i2}b_{2r}c_{rj}\\
&\quad +\ldots\\
&\quad +a_{in}b_{n1}c_{1j}+a_{in}b_{n2}c_{2j}+\ldots+a_{in}b_{nr}c_{rj}\\
&= \left(a_{i1}b_{11}+a_{i2}b_{21}+\ldots+a_{in}b_{n1}\right)c_{1j}\\
&\quad +\left(a_{i1}b_{12}+a_{i2}b_{22}+\ldots+a_{in}b_{n2}\right)c_{2j}\\
&\quad +\ldots\\
&\quad +\left(a_{i1}b_{1r}+a_{i2}b_{2r}+\ldots+a_{in}b_{nr}\right)c_{rj}\\
&= \left(\mathbf{AB}\right)_{i1}c_{1j}+\left(\mathbf{AB}\right)_{i2}c_{2j}+\ldots+\left(\mathbf{AB}\right)_{ir}c_{rj}\\
&= \left[\left(\mathbf{AB}\right)\mathbf{C}\right]_{ij}
\end{aligned}
$$

Since both matrices have the same size, and their corresponding entries are equal, $\mathbf{A}\left(\mathbf{BC}\right)=\left(\mathbf{AB}\right)\mathbf{C}$.

Associate law for matrix multiplication allows us to write **ABC** without ambiguity if **A**, **B** and **C** are matrices of appropriate sizes.

The commutative law for matrix, $\mathbf{AB}=\mathbf{BA}$, is obviously <u>not true</u> if **A** is of size $m\times n$, **B** is of size $n\times m$ and $m\neq n$, as the products are matrices of different sizes.

**Example 2.3.1**

Prove or disprove the statement: $\mathbf{AB}=\mathbf{BA}$ for any matrices **A** and **B** of the same size $n\times n$.

*Solution:*

The statement is false. (We just need to provide a counterexample)

Let $\mathbf{A}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$ and $\mathbf{B}=\begin{pmatrix}0&1\\0&0\end{pmatrix}$, then $\mathbf{AB}=\begin{pmatrix}0&1\\0&0\end{pmatrix}$ and $\mathbf{BA}=\begin{pmatrix}0&0\\0&0\end{pmatrix}$, $\mathbf{AB}\neq\mathbf{BA}$.

The above example shows that in general, matrix multiplication is not commutative, that is, **AB** need not be equal to **BA**, even if both **AB** and **BA** are defined and of the same size.

**Theorem 2.3.3**

Let $r$ and $s$ be real numbers and let **A** and **B** be matrices of appropriate sizes so that the indicated operations are defined, then

$$r(s\mathbf{A}) = (rs)\mathbf{A}$$
$$(r+s)\mathbf{A} = r\mathbf{A} + s\mathbf{A}$$
$$r(\mathbf{A}+\mathbf{B}) = r\mathbf{A} + r\mathbf{B}$$
$$r(\mathbf{AB}) = (r\mathbf{A})\mathbf{B} = \mathbf{A}(r\mathbf{B})$$

**Theorem 2.3.4**

Let $k$ be a real number and let **A** and **B** be matrices of appropriate sizes so that the indicated operations are defined, then

$$\left(\mathbf{A}^T\right)^T = \mathbf{A}$$
$$(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$$
$$(k\mathbf{A})^T = k\left(\mathbf{A}^T\right)$$
$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$$

*Proof* (for $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$):

Let **A** and **B** be $m \times n$ and $n \times r$ matrices respectively.

First note that **AB** is of size $m \times r$, thus $(\mathbf{AB})^T$ is of size $r \times m$. $\mathbf{A}^T$ and $\mathbf{B}^T$ are of sizes $n \times m$ and $r \times n$ respectively, thus $\mathbf{B}^T\mathbf{A}^T$ is of size $r \times m$. Therefore $(\mathbf{AB})^T$ and $\mathbf{B}^T\mathbf{A}^T$ have the same size.

For any $i, j$ where $1 \le i \le r$ and $1 \le j \le m$, we have

$$\left[(\mathbf{AB})^T\right]_{ij} = (\mathbf{AB})_{ji} = a_{j1}b_{1i} + a_{j2}b_{2i} + \ldots + a_{jn}b_{ni}.$$

On the other hand,

$$\begin{aligned}\left(\mathbf{B}^T\mathbf{A}^T\right)_{ij} &= \left(\mathbf{B}^T\right)_{i1}\left(\mathbf{A}^T\right)_{1j} + \left(\mathbf{B}^T\right)_{i2}\left(\mathbf{A}^T\right)_{2j} + \ldots\left(\mathbf{B}^T\right)_{in}\left(\mathbf{A}^T\right)_{nj} \\ &= (\mathbf{B})_{1i}(A)_{j1} + (\mathbf{B})_{2i}(A)_{j2} + \ldots + (\mathbf{B})_{ni}(A)_{jn} \\ &= b_{1i}a_{j1} + b_{2i}a_{j2} + \ldots + b_{ni}a_{jn} \\ &= a_{j1}b_{1i} + a_{j2}b_{2i} + \ldots + a_{jn}b_{ni}.\end{aligned}$$

We see that $\left[(\mathbf{AB})^T\right]_{ij} = \left(\mathbf{B}^T\mathbf{A}^T\right)_{ij}$. Hence $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

**Zero Matrices**

We know that the real number 0 has the special property that for any real number $a$, we have

$$a+0=0+a=a\ .$$

We have matrices that play similar role as that of 0 for real numbers.

> **Definition**
>
> A matrix of all whose entries are zero is called a ***zero matrix***.

For example, $\begin{pmatrix}0&0\\0&0\end{pmatrix}$, $\begin{pmatrix}0&0&0\\0&0&0\end{pmatrix}$, $\begin{pmatrix}0&0\end{pmatrix}$ and $\begin{pmatrix}0\\0\\0\end{pmatrix}$ are all zero matrices.

A zero matrix is denoted by **O**. If it is important to emphasize the size, we shall write $\mathbf{O}_{m\times n}$ for the $m\times n$ zero matrix.

> **Theorem 2.3.5**
>
> Assume the matrices are of appropriate sizes such that the indicated operations are defined, then
>
> $\mathbf{A}+\mathbf{O}=\mathbf{O}+\mathbf{A}=\mathbf{A}$
> $\mathbf{A}-\mathbf{A}=\mathbf{O}$
> $\mathbf{AO}=\mathbf{O}$ and $\mathbf{OA}=\mathbf{O}$

**Example 2.3.2**

Prove or disprove the statement: if $\mathbf{AB}=\mathbf{O}$, then $\mathbf{A}=\mathbf{O}$ or $\mathbf{B}=\mathbf{O}$.

*Solution:*

The statement is false. A counterexample can be

let $\mathbf{A}=\begin{pmatrix}1&2\\2&4\end{pmatrix}$ and $\mathbf{B}=\begin{pmatrix}4&-6\\-2&3\end{pmatrix}$, $\mathbf{AB}=\begin{pmatrix}0&0\\0&0\end{pmatrix}=\mathbf{O}$ but neither matrix is **O**.

**Example 2.3.3**

Prove or disprove the statement: if $\mathbf{AB}=\mathbf{AC}$ and $\mathbf{A}\neq\mathbf{O}$, then $\mathbf{B}=\mathbf{C}$ .

*Solution:*

The statement is false. For example,
let $\mathbf{A}=\begin{pmatrix}1&2\\2&4\end{pmatrix}$, $\mathbf{B}=\begin{pmatrix}2&-3\\1&2\end{pmatrix}$ and $\mathbf{C}=\begin{pmatrix}-2&3\\3&-1\end{pmatrix}$, $\mathbf{AB}=\begin{pmatrix}4&1\\8&2\end{pmatrix}=\mathbf{AC}$.

## Identity Matrices

For real numbers, the number 1 has the special property that $1\times a = a\times 1 = a$ for all numbers $a$. For matrices, we also have matrices that have similar property.

> **Definition**
>
> A <u>square</u> matrix with '1's on the main diagonal and 0's off the main diagonal is called an ***identity matrix***.

For example, $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ are identity matrices.

An identity matrix is denoted by $\mathbf{I}$. If it is important to emphasize the size, we shall write $\mathbf{I}_n$ for the $n\times n$ identity matrix.

### Example 2.3.4

Let $\mathbf{A} = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$, evaluate $\mathbf{AI}_3$ and $\mathbf{I}_2\mathbf{A}$.

*Solution:*

$\mathbf{AI}_3 = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$ and $\mathbf{I}_2\mathbf{A} = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$

> **Theorem 2.3.6**
>
> If $\mathbf{A}$ is an $m\times n$ matrix, then $\mathbf{AI}_n = \mathbf{I}_m\mathbf{A} = \mathbf{A}$. In particular, for any $n\times n$ square matrix $\mathbf{B}$, we have $\mathbf{BI}_n = \mathbf{I}_n\mathbf{B} = \mathbf{B}$.

You may prove this theorem as a practice.

**Example 2.3.5**

Prove or disprove the following statements.

**(a)** If $\mathbf{AB} = \mathbf{BA} = \mathbf{A}$ for some nonzero $n \times n$ matrice **A**, then $\mathbf{B} = \mathbf{I}$.
**(b)** If $\mathbf{AC} = \mathbf{A}$ for all $n \times n$ matrices **A**, then $\mathbf{C} = \mathbf{I}$.
**(c)** If $\mathbf{AD} = \mathbf{DA}$ for all $n \times n$ matrices **A**, then $\mathbf{D} = \mathbf{I}$ or $\mathbf{D} = \mathbf{O}$.

*Solution:*

**(a)** The statement is false.

A counterexample: let $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

$\mathbf{AB} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{BA} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, so $\mathbf{AB} = \mathbf{BA} = \mathbf{A}$.

Here $\mathbf{B} \neq \mathbf{I}$.

**(b)** The statement is true. (We suspect this after failing to find a counterexample after numerous attempt)

Here is a demonstration for $n = 2$, you may try to prove the statement for a general $n$.

Let $\mathbf{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$.

When $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $\mathbf{AC} = \begin{pmatrix} c_{11} & c_{12} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, so $c_{11} = 1$ and $c_{12} = 0$.

When $\mathbf{A} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{AC} = \begin{pmatrix} 0 & 0 \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, so $c_{21} = 0$ and $c_{22} = 1$.

When $\mathbf{C} = \mathbf{I}$, $\mathbf{AC} = \mathbf{A}$ for all $n \times n$ matrices A by **Theorem 2.3.6**.

Therefore the statement is true.

**(c)** The statement is not true. All matrices $\mathbf{D} = k\mathbf{I}$ where $k \in \mathbb{R}$ obviously has this property.

- Considering the proof for **(b)**, is it true that if $\mathbf{UA} = \mathbf{A}$ for all $n \times n$ matrices **A**, then $\mathbf{U} = \mathbf{I}$?
- In **(c)**, can we find other matrices apart from $k\mathbf{I}$ that have this property?

**Definition**

Let $\mathbf{A}$ be a square matrix. If $p$ is a positive integer, we define $\mathbf{A}^p = \underbrace{\mathbf{AA}\ldots\mathbf{A}}_{p \text{ factors}}$. We also define $\mathbf{A}^0 = \mathbf{I}$.

**Example 2.3.6**

Prove or disprove the statement: if $k > 1$ be a positive integer, $(\mathbf{AB})^k = \mathbf{A}^k\mathbf{B}^k$ for all $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$.

*Solution:*

The statement is false.

A counterexample: let $k = 2$, $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

$\mathbf{AB} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, $(\mathbf{AB})^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$. $\mathbf{A}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{B}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\mathbf{A}^2\mathbf{B}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Here $(\mathbf{AB})^k \neq \mathbf{A}^k\mathbf{B}^k$.

**Theorem 2.3.7**

If $\mathbf{A}$ is a square matrix and $r$ and $s$ are nonnegative integers, then

$$\mathbf{A}^r\mathbf{A}^s = \mathbf{A}^{r+s}$$

$$\left(\mathbf{A}^r\right)^s = \mathbf{A}^{rs}$$

## §3 Inverse Matrix and Its Applications

### 3.1 Inverse of a Matrix

For any nonzero real number *a*, we can find a real number *b* such that $ab = ba = 1$.

Since identity matrices play similar role for matrices as 1 for real numbers with respect to multiplication, it is natural to ask the following question: given any nonzero $n \times n$ square matrix **A**, can we find an $n \times n$ matrix **B** such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$?

**Definition**

Let **A** be an $n \times n$ square matrix. If there exists an $n \times n$ matrix **B** such that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n,$$

then we say that **A** is ***invertible*** or ***nonsingular***, and in this case, **B** is called an ***inverse*** of **A**. If no such matrix **B** exists, the we say that **A** is ***noninvertible*** or ***singular***.

For example, let $\mathbf{A} = \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix}$. Then $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix}$ is an inverse of **A** since $\mathbf{AB} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}$ and $\mathbf{BA} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}$.

**Theorem 3.1.1**

If a matrix **A** is invertible, then its inverse is unique.

*Proof:*

Let **B** and **C** be inverses of **A**. Then $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$ and $\mathbf{AC} = \mathbf{CA} = \mathbf{I}$. We have

$$\mathbf{B} = \mathbf{IB} = (\mathbf{CA})\mathbf{B} = \mathbf{C}(\mathbf{AB}) = \mathbf{CI} = \mathbf{C}.$$

Hence **A** has a unique inverse.

In view of **Theorem 3.1.1**, we shall now speak of 'the' inverse of an invertible matrix.

**Notation:** If **A** be an invertible matrix, then the inverse of **A** is denoted by $\mathbf{A}^{-1}$.

- Given a square matrix **A**, how do we determine whether it is invertible?
- If **A** is invertible, how do we find its inverse?

**Theorem 3.1.2**

Let $\mathbf{A}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $ad-bc\neq 0$, then $\mathbf{A}$ is invertible, and its inverse is given by

$$\mathbf{A}^{-1}=\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

*Proof:*

Just verify that the given formula for $\mathbf{A}^{-1}$ satisfies the definition, i.e. $\mathbf{A}\mathbf{A}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$.

**Theorem 3.1.3**

**(a)** If $\mathbf{A}$ is an invertible matrix, then $\mathbf{A}^{-1}$ is invertible and

$$\left(\mathbf{A}^{-1}\right)^{-1}=\mathbf{A}.$$

**(b)** If $\mathbf{A}$ and $\mathbf{B}$ are invertible matrices of the same size, then $\mathbf{AB}$ is invertible and

$$\left(\mathbf{AB}\right)^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}. \text{ (Sock-Shoes rule)}$$

**(c)** If $\mathbf{A}$ is invertible, then $k\mathbf{A}$ is invertible for any nonzero scalar $k$, and

$$\left(k\mathbf{A}\right)^{-1}=\frac{1}{k}\mathbf{A}^{-1}.$$

**(d)** If $\mathbf{A}$ is invertible, then $\mathbf{A}^T$ is invertible and

$$\left(\mathbf{A}^T\right)^{-1}=\left(\mathbf{A}^{-1}\right)^T.$$

**(e)** If $\mathbf{A}$ is invertible, then it cannot a row or a column of zeros.

*Proof:*

**(a)** Since $\mathbf{A}\mathbf{A}^{-1}=\mathbf{A}^{-1}\mathbf{A}=\mathbf{I}$, this shows that $\mathbf{A}^{-1}$ is invertible and the inverse of $\mathbf{A}^{-1}$ is $\mathbf{A}$, that is $\left(\mathbf{A}^{-1}\right)^{-1}=\mathbf{A}$.

**(b)** $\left(\mathbf{AB}\right)\left(\mathbf{B}^{-1}\mathbf{A}^{-1}\right)=\mathbf{ABB}^{-1}\mathbf{A}=\mathbf{A}\left(\mathbf{BB}^{-1}\right)\mathbf{A}^{-1}=\mathbf{AIA}^{-1}=\mathbf{AA}^{-1}=\mathbf{I}$, and similarly, $\left(\mathbf{B}^{-1}\mathbf{A}^{-1}\right)\left(\mathbf{AB}\right)=\mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB}=\mathbf{B}^{-1}\left(\mathbf{A}^{-1}\mathbf{A}\right)\mathbf{B}=\mathbf{B}^{-1}\mathbf{IB}=\mathbf{B}^{-1}\mathbf{B}=\mathbf{I}$. Hence $\mathbf{B}^{-1}\mathbf{A}^{-1}$ is the inverse of $\mathbf{AB}$, i.e. $\left(\mathbf{AB}\right)^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}$.

- If $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are invertible matrices of the same size, then is $\mathbf{ABC}$ invertible?

The proofs for **(c)** and **(d)** are similar.

**(e)** Suppose row $i$ contains only 0s, i.e. $\left(\mathbf{A}\right)_{ik}=0$ for all $1\le k\le n$. Then for any $n\times n$ matrix $\mathbf{B}$, the $(i,i)$ entry in the product $\left(\mathbf{AB}\right)_{ii}=\left(\mathbf{A}\right)_{i1}\left(\mathbf{A}^{-1}\right)_{1i}+\left(\mathbf{A}\right)_{i2}\left(\mathbf{A}^{-1}\right)_{2i}+\ldots+\left(\mathbf{A}\right)_{in}\left(\mathbf{A}^{-1}\right)_{ni}=0$. However, this entry should be 1 if $\mathbf{B}$ is to be the inverse of $\mathbf{A}$. Thus $\mathbf{A}$ has no inverse.

**Example 3.1.1**

Suppose **A** and **B** are matrices such that $\mathbf{A}^{-1}=\begin{pmatrix}1&2\\0&1\end{pmatrix}$ and $\mathbf{B}^{-1}=\begin{pmatrix}2&1\\3&2\end{pmatrix}$, find the inverse of **AB**.

*Solution:*

By **Theorem 3.1.3(b)**, $(\mathbf{AB})^{-1}=\mathbf{B}^{-1}\mathbf{A}^{-1}=\begin{pmatrix}2&1\\3&2\end{pmatrix}\begin{pmatrix}1&2\\0&1\end{pmatrix}=\begin{pmatrix}2&5\\3&8\end{pmatrix}$.

**Example 3.1.2**

Suppose **A** is a $3\times3$ matrix such that $\mathbf{A}^{-1}=\begin{pmatrix}-40&16&9\\13&-5&-3\\5&-2&-1\end{pmatrix}$. Find the inverses of 2**A** and $\mathbf{A}^T$.

*Solution:*

By **Theorem 3.1.3(c)**, $(2\mathbf{A})^{-1}=\frac{1}{2}\mathbf{A}^{-1}=\begin{pmatrix}-20&8&4.5\\6.5&-2.5&-1.5\\2.5&-1&-0.5\end{pmatrix}$.

**By Theorem 3.1.3(d)**, $(\mathbf{A}^T)^{-1}=(\mathbf{A}^{-1})^T=\begin{pmatrix}-40&13&5\\16&-5&-2\\9&-3&-1\end{pmatrix}$.

Recall that we have defined the powers of a matrix $\mathbf{A}^n$ for nonnegative integer *n*. We can extend the definition to negative integer powers if the matrix is <u>invertible</u>.

**Definition**

Let **A** be an <u>invertible</u> matrix and let *n* be a positive integer. Then we define

$$\mathbf{A}^{-n}=(\mathbf{A}^{-1})^n=\underbrace{\mathbf{A}^{-1}\mathbf{A}^{-1}...\mathbf{A}^{-1}}_{n\text{ factors}}.$$

For example, $\mathbf{A}^{-3}=\mathbf{A}^{-1}\mathbf{A}^{-1}\mathbf{A}^{-1}$.

**Theorem 3.1.4** (Comparing to **Theorem 2.3.7**)

If **A** is an invertible matrix and *r* and *s* are integers, then

$$\mathbf{A}^r\mathbf{A}^s=\mathbf{A}^{r+s}$$
$$(\mathbf{A}^r)^s=\mathbf{A}^{rs}$$

- Is $\mathbf{A}^{-3}=(\mathbf{A}^3)^{-1}$, where **A** is an invertible matrix?

## 3.2 Elementary Matrices

**Definition**

An $n \times n$ square matrix is called an ***elementary matrix*** if it can be obtained from the $n \times n$ identity matrix by performing a single elementary row operation (recall its definition in **Section 1.2**).

**Example 3.2.1**

Determine whether each of the matrices below is an elementary matrix.

**(a)** $\begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$, **(b)** $\begin{pmatrix} 1 & 0 \\ 0 & -3 \end{pmatrix}$, **(c)** $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$, **(d)** $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$, **(e)** $\begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, **(f)** $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

*Solution:*

**(a)**, **(d)** and **(f)** are not elementary matrices.
**(b)** can be obtained by multiplying row 2 by $-3$ in $\mathbf{I}_2$.
**(c)** can be obtained by exchanging row 1 and row 3 in $\mathbf{I}_3$.
**(e)** can be obtained by adding 5 times row 3 to row 1 in $\mathbf{I}_3$.
Thus, **(b)**, **(c)** and **(e)** are elementary matrices.

**Example 3.2.2**

Consider a general $3 \times 4$ matrix

$$\mathbf{A} = \begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \end{pmatrix},$$

and three elementary matrices

$$\mathbf{E}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \ \mathbf{E}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{E}_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{pmatrix}.$$

**(i)** Find $\mathbf{E}_1\mathbf{A}$, $\mathbf{E}_2\mathbf{A}$ and $\mathbf{E}_3\mathbf{A}$.

**(ii)** Determine whether the results of **(i)** can be obtained by performing a certain elementary row operation on **A** respectively.

**(iii)** Considering the respective elementary row operations to be performed on **I** to obtain $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$, what is the significance of the results of **(ii)**?

*Solution:*

$\mathbf{E}_1\mathbf{A}=\begin{pmatrix} i & j & k & l \\ e & f & g & h \\ a & b & c & d \end{pmatrix}$, which can be obtained by interchanging row 1 and row 3 in **A**.

$\mathbf{E}_2\mathbf{A}=\begin{pmatrix} a & b & c & d \\ 4e & 4f & 4g & 4h \\ i & j & k & l \end{pmatrix}$, which can be obtained by multiplying row 2 by 4 in **A**.

$\mathbf{E}_3\mathbf{A}=\begin{pmatrix} a & b & c & d \\ e & f & g & h \\ i+3a & j+3b & k+3c & l+4d \end{pmatrix}$, which can be obtained by adding 3 × row 1 to row 3 in **A**.

These elementary row operations are the same as those to be performed on **I** to obtain $\mathbf{E}_1$, $\mathbf{E}_2$ and $\mathbf{E}_3$ respectively.

**Theorem 3.2.1**

If the elementary matrix **E** results from performing a certain row operation on $\mathbf{I}_m$ and if **A** is an $m\times n$ matrix, then the product **EA** is the matrix that results when this same row operation is performed on **A**.

The above theorem is illustrated by the following diagram, where *r* denotes an elementary row operation:

(1) $\mathbf{A}\xrightarrow{r}\mathbf{B}$
(2) $\mathbf{I}_m\xrightarrow{r}\mathbf{E}$
(3) $\mathbf{EA}=\mathbf{B}$

Given (2), (3) implies (1). Given (2), (1) implies (3).

**Example 3.2.3**

Consider the $3\times 4$ matrix $\mathbf{A}=\begin{pmatrix} 0 & 1 & 2 & 1 \\ 0 & 4 & 0 & 1 \\ 2 & -2 & 6 & 4 \end{pmatrix}$, and the elementary matrices

$$\mathbf{E}_1=\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},\ \mathbf{E}_2=\begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } \mathbf{E}_3=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}.$$

Find $\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A}$ and $\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3\mathbf{A}$.

*Solution:*

The elementary row operations are
$r_1$: interchanging row 1 and row 3, $r_2$: multiplying row 1 by 0.5, $r_3$: adding $-2$ times row 2 to row 3.

$$\mathbf{A}\xrightarrow{r_1}\mathbf{E}_1\mathbf{A}\xrightarrow{r_2}\mathbf{E}_2\left(\mathbf{E}_1\mathbf{A}\right)\xrightarrow{r_3}\mathbf{E}_3\left[\mathbf{E}_2\left(\mathbf{E}_1\mathbf{A}\right)\right]=\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A}$$

$$\begin{pmatrix}0&1&2&1\\0&4&0&1\\2&-2&6&4\end{pmatrix}\xrightarrow{r_1}\begin{pmatrix}2&-2&6&4\\0&4&0&1\\0&1&2&1\end{pmatrix}\xrightarrow{r_2}\begin{pmatrix}1&-1&3&2\\0&4&0&1\\0&1&2&1\end{pmatrix}\xrightarrow{r_3}\begin{pmatrix}1&-1&3&2\\0&4&0&1\\0&-7&2&-1\end{pmatrix}=\mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A}$$

$$\mathbf{A}\xrightarrow{r_3}\mathbf{E}_3\mathbf{A}\xrightarrow{r_2}\mathbf{E}_2\left(\mathbf{E}_3\mathbf{A}\right)\xrightarrow{r_1}\mathbf{E}_1\left[\mathbf{E}_2\left(\mathbf{E}_3\mathbf{A}\right)\right]=\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3\mathbf{A}$$

$$\begin{pmatrix}0&1&2&1\\0&4&0&1\\2&-2&6&4\end{pmatrix}\xrightarrow{r_3}\begin{pmatrix}0&1&2&1\\0&4&0&1\\2&-10&6&2\end{pmatrix}\xrightarrow{r_2}\begin{pmatrix}0&0.5&1&0.5\\0&4&0&1\\2&-10&6&2\end{pmatrix}\xrightarrow{r_3}\begin{pmatrix}2&-10&6&2\\0&4&0&1\\0&0.5&1&0.5\end{pmatrix}=\mathbf{E}_1\mathbf{E}_2\mathbf{E}_3\mathbf{A}$$

**Theorem 3.2.2**

Every elementary matrix is invertible, and the inverse is also an elementary matrix.

**Theorem 3.2.3**

Let **A** be a square matrix. Then **A** is invertible if and only if its reduced row-echelon form (recall its definition in **Section 1.2**) is the identity matrix.

*Proof:*

Let **R** be the reduced row-echelon form of **A**. Thus **R** is obtained by performing a sequence of elementary row operations on **A**.

By **Theorem 3.2.1**, there exist elementary matrices $\mathbf{E}_1$, $\mathbf{E}_2$, …, $\mathbf{E}_k$ such that $\mathbf{E}_k...\mathbf{E}_2\mathbf{E}_1\mathbf{A}=\mathbf{R}$.
Let $\mathbf{B}=\mathbf{E}_k...\mathbf{E}_2\mathbf{E}_1$, we have $\mathbf{BA}=\mathbf{R}$.

Since every elementary matrix is invertible by **Theorem 3.2.2**, their product, **B**, is invertible by **Theorem 3.1.3(b)**.

(To prove 'if') Suppose the reduced row-echelon form of **A** is **I**, i.e. $\mathbf{BA}=\mathbf{R}=\mathbf{I}$.
Since **B** is invertible, $\mathbf{B}^{-1}\left(\mathbf{BA}\right)=\mathbf{B}^{-1}\mathbf{I}$. We know $\mathbf{B}^{-1}\left(\mathbf{BA}\right)=\left(\mathbf{B}^{-1}\mathbf{B}\right)\mathbf{A}=\mathbf{IA}=\mathbf{A}$ and $\mathbf{B}^{-1}\mathbf{I}=\mathbf{B}^{-1}$, so $\mathbf{A}=\mathbf{B}^{-1}$ which is invertible by **Theorem 3.1.3(a)**.

(To prove 'only if') Now suppose that **A** is invertible of size $n\times n$.
Since both **B** and **A** are invertible, their product, **R**, is also invertible by **Theorem 3.2.3(b)**. Therefore **R**, a $n\times n$ matrix in reduced row-echelon form, cannot have a row of zeros by **Theorem 3.2.3(e)**. Thus **R** must have exactly $n$ 1s in total. Because these leading 1's occur progressively to the right as we move down the rows and the last row must still contain a 1, they must be on the main diagonal. Since all the other entries are 0s, **R** must be the identity matrix.

### 3.3 A Method for Finding Inverse

Suppose **A** is an invertible matrix. Then by **Theorem 3.2.3**, we can perform a sequence of elementary row operations on **A** to produce **I**. By **Theorem 3.2.1**, we can find elementary matrices $\mathbf{E}_1$, $\mathbf{E}_2$, …, $\mathbf{E}_k$ such that $\mathbf{E}_k ... \mathbf{E}_2\mathbf{E}_1\mathbf{A} = \mathbf{I}$.

Multiplying both sides on the right by $\mathbf{A}^{-1}$, we obtain $\mathbf{E}_k ... \mathbf{E}_2\mathbf{E}_1\mathbf{I} = \mathbf{A}^{-1}$.

This result gives us an algorithm for finding the inverse of an invertible matrix: perform a sequence of elementary row operations on **A** to reduce it to **I**, then perform the same sequence of elementary row operations on **I** to obtain $\mathbf{A}^{-1}$.

**Example 3.3.1**

Find the inverse of $\mathbf{A} = \begin{pmatrix} 1 & -1 & 2 \\ -5 & 7 & -11 \\ -2 & 3 & -5 \end{pmatrix}$.

*Solution:*

We form the portioned matrix $(\mathbf{A} \mid \mathbf{I})$ by adjoining the identity matrix to the right of **A**, then perform elementary row operations to the matrix till the left side is reduced to **I**, and the right side will be $\mathbf{A}^{-1}$.

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ -5 & 7 & -11 & 0 & 1 & 0 \\ -2 & 3 & -5 & 0 & 0 & 1 \end{array}\right) \to\to \left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 2 & -1 & 5 & 1 & 0 \\ 0 & 1 & -1 & 2 & 0 & 1 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & 2 & 0 & 1 \\ 0 & 2 & -1 & 5 & 1 & 0 \end{array}\right) \to$$

$$\left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 1 & -1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & -2 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 3 & 1 & -1 \\ 0 & 0 & 1 & 1 & 1 & -2 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & -1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 3 & 1 & -1 \\ 0 & 0 & 1 & 1 & 1 & -2 \end{array}\right) \to$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 4 & 1 & -1 \\ 0 & 1 & 0 & 3 & 1 & -1 \\ 0 & 0 & 1 & 1 & 1 & -2 \end{array}\right) \to \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & -1 & 3 \\ 0 & 1 & 0 & 3 & 1 & -1 \\ 0 & 0 & 1 & 1 & 1 & -2 \end{array}\right)$$

Thus $\mathbf{A}^{-1} = \begin{pmatrix} 2 & -1 & 3 \\ 3 & 1 & -1 \\ 1 & 1 & -2 \end{pmatrix}$.

- What will happen if we use the algorithm on a noninvertible square matrix?

### 3.4 Results on Linear System and Invertibility

We have seen in **Section 2.2** that every linear system can be written as a matrix equation $\mathbf{Ax} = \mathbf{b}$. Using this matrix equation and the properties of matrix operations, we are able to prove **Theorem 1.1.1**.

> **Theorem 1.1.1**
>
> Every system of linear equations has either no solution, exactly one solution or infinitely many solutions. (There are no other possibilities)

*Proof:*

Let $\mathbf{Ax} = \mathbf{b}$ … (1) be a linear system. If the linear system has no solution or exactly one solution (which can happen), then we have completed the proof.

Now assume that the linear system (1) has more than one solution, then we want to show that it has infinitely many solutions. Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two distinct solutions of (1), and let $\mathbf{x}_0 = \mathbf{x}_1 - \mathbf{x}_2$, then $\mathbf{x}_0 \neq \mathbf{0}$, and we have

$$\mathbf{Ax}_0 = \mathbf{A}\left(\mathbf{x}_1 - \mathbf{x}_2\right) = \mathbf{Ax}_1 - \mathbf{Ax}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

Now for any real scalar *k*,

$$\mathbf{A}\left(\mathbf{x}_1 + k\mathbf{x}_0\right) = \mathbf{Ax}_1 + \mathbf{A}\left(k\mathbf{x}_0\right) = \mathbf{Ax}_1 + k\left(\mathbf{Ax}_0\right) = \mathbf{b} + k\mathbf{0} = \mathbf{b}.$$

This shows that $\mathbf{x}_1 + k\mathbf{x}_0$ is also a solution of (1). Since $\mathbf{x}_0 \neq \mathbf{0}$ and there are infinitely many values for *k*, we conclude that (1) now has infinitely many solutions. (Does this idea sound look familiar?)

> **Theorem 3.4.1**
>
> If $\mathbf{A}$ is an <u>invertible</u> $n \times n$ matrix, then for any $n \times 1$ matrix **b**, the system of linear equations $\mathbf{Ax} = \mathbf{b}$ has <u>exactly one</u> solution, namely $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

**Example 3.4.1**

Find the solution of the following linear system using **Theorem 3.4.1**.

$$\begin{aligned} 4x \quad -3y &= -3 \\ 2x \quad -5y &= 9 \end{aligned}$$

*Solution:*

$\mathbf{A} = \begin{pmatrix} 4 & -3 \\ 2 & -5 \end{pmatrix}$. To find $\mathbf{A}^{-1}$:

$$\left(\begin{array}{cc|cc} 4 & -3 & 1 & 0 \\ 2 & -5 & 0 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & -\frac{3}{4} & \frac{1}{4} & 0 \\ 2 & -5 & 0 & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & -\frac{3}{4} & \frac{1}{4} & 0 \\ 0 & -\frac{7}{2} & -\frac{1}{2} & 1 \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & -\frac{3}{4} & \frac{1}{4} & 0 \\ 0 & 1 & \frac{1}{7} & -\frac{2}{7} \end{array}\right) \to \left(\begin{array}{cc|cc} 1 & 0 & \frac{5}{14} & -\frac{3}{14} \\ 0 & 1 & \frac{1}{7} & -\frac{2}{7} \end{array}\right)$$

$\mathbf{A}^{-1} = \begin{pmatrix} \frac{5}{14} & -\frac{3}{14} \\ \frac{1}{7} & -\frac{2}{7} \end{pmatrix}$, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \begin{pmatrix} \frac{5}{14} & -\frac{3}{14} \\ \frac{1}{7} & -\frac{2}{7} \end{pmatrix}\begin{pmatrix} -3 \\ 9 \end{pmatrix} = \begin{pmatrix} -3 \\ -3 \end{pmatrix}$.

Thus $x = -3$ and $y = -3$.

**Theorem 3.4.2 (**Compare with the definition in **Section 3.1**)

Let **A** be an $n \times n$ matrix.

**(a)** If there exists an $n \times n$ matrix **B** such that $\mathbf{BA} = \mathbf{I}$, then **A** is invertible and $\mathbf{B} = \mathbf{A}^{-1}$.
**(b)** If there exists an $n \times n$ matrix **B** such that $\mathbf{AB} = \mathbf{I}$, then **A** is invertible and $\mathbf{B} = \mathbf{A}^{-1}$.

*Proof:*

**(a)** We can prove **A** is invertible by **Theorem 3.2.3**.

Consider the homogeneous linear system $\mathbf{Ax} = \mathbf{0}$.

Multiplying both sides of the equation on the left by **B**, we obtain $\mathbf{BAx} = \mathbf{B0}$. This gives $\mathbf{Ix} = \mathbf{0}$, i.e. $\mathbf{x} = \mathbf{0}$, the trivial solution. This implies that if we solve the homogeneous system by Gauss-Jordan elimination on the augmented matrix $(\mathbf{A} \mid \mathbf{0})$ by reducing it to reduced row-echelon form, we would get $(\mathbf{I} \mid \mathbf{0})$. Consequently, the reduced row-echelon form of **A** is the identity matrix. Hence **A** is invertible by **Theorem 3.2.3**.

Since **A** is invertible and $\mathbf{BA} = \mathbf{I}$, we have $\mathbf{BAA}^{-1} = \mathbf{IA}^{-1}$, which gives $\mathbf{BI} = \mathbf{A}^{-1}$, i.e. $\mathbf{B} = \mathbf{A}^{-1}$.

**(b)** Take the transpose on both sides of $\mathbf{AB} = \mathbf{I}$.

$$(\mathbf{AB})^T = \mathbf{I}^T$$
$$\mathbf{B}^T\mathbf{A}^T = \mathbf{I}$$

By part **(a)**, $\mathbf{A}^T$ is invertible and its inverse is $\mathbf{B}^T$.

Since $\mathbf{A} = \left(\mathbf{A}^T\right)^T$, by Theorem **3.1.3(d)**, **A** is also invertible and its inverse is $\left(\mathbf{B}^T\right)^T$, which is **B**.

**Theorem 3.4.3**

Let **A** be an $n \times n$ matrix. Then the following statements are equivalent:

(1) **A** is invertible.
(2) The linear system $\mathbf{Ax} = \mathbf{0}$ has <u>only</u> the trivial solution, i.e. $\mathbf{x} = \mathbf{0}$ is the only solution.
(3) The reduced row-echelon form of **A** is **I**.
(4) **A** can be expressed as a product of elementary matrices.
(5) $\mathbf{Ax} = \mathbf{b}$ is consistent for <u>every</u> $n \times 1$ matrix **b**.
(6) $\mathbf{Ax} = \mathbf{b}$ has <u>exactly one</u> solution for <u>every</u> $n \times 1$ matrix **b**.

# §4 Determinants

## 4.1 Determinants by Cofactor Expansions

Recall the a $2\times 2$ matrix $\mathbf{A}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if $ad-bc\neq 0$. The number $ad-bc$ is called he determinant of **A**, and is noted by $\det(\mathbf{A})$ or $|\mathbf{A}|$. Prior to defining the determinant of an $n\times n$ matrix, we need to define a few relevant quantities first.

**Definition**

Suppose we have defined the determinant of $(n-1)\times(n-1)$ matrix, for $n\geq 2$

Let $\mathbf{A}=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$ …(1) be an $n\times n$ matrix for $n\geq 2$.

Let $M_{ij}$ be the determinant of the $(n-1)\times(n-1)$ submatrix obtained from **A** by deleting the row and the column that contain $a_{ij}$, i.e. the *i*th row and *j*th column of **A**. The number $M_{ij}$ is called the ***minor*** of the entry $a_{ij}$. The ***cofactor*** of entry $a_{ij}$ is defined to be the number $(-1)^{i+j}M_{ij}$, and is denoted by $C_{ij}$.

**Example 4.1.1**

Let $\mathbf{A}=\begin{pmatrix} 5 & 0 & 8 \\ 2 & -1 & 3 \\ 4 & 1 & 0 \end{pmatrix}$. Find $M_{11}$, $C_{11}$, $M_{32}$ and $C_{32}$.

*Solution:*

$M_{11}=\begin{vmatrix} -1 & 3 \\ 1 & 0 \end{vmatrix}=-3$, $C_{11}=(-1)^{1+1}(-3)=-3$. $M_{32}=\begin{vmatrix} 5 & 8 \\ 2 & 3 \end{vmatrix}=-1$, $C_{32}=(-1)^{3+2}(-1)=1$.

**Definition**

Let **A** be an $n \times n$ matrix in (1).

The ***cofactor expansion of A along row i***, $1 \le i \le n$, is the expression

$$\sum_{j=1}^{n} a_{ij}C_{ij} = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in} \ \ldots(2)$$

The ***cofactor expansion of A column row j***, $1 \le j \le n$, is the expression

$$\sum_{i=1}^{n} a_{ij}C_{ij} = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj} \ \ldots \ (3)$$

**Theorem 4.1.1**

Let **A** be an $n \times n$ matrix in (1). The values given by expressions (2) and (3) are equal, regardless of the row or column chosen.

Now we are ready to define the determinant of an $n \times n$ matrix <u>inductively</u>.

**Definition**

The ***determinant*** of a $1\times1$ matrix, $(a)$, is $a$.

Let **A** be an $n \times n$ matrix in (1) for $n \ge 2$. Then we defined the common value in (2) and (3) to be the ***determinant*** of **A**, and denote by $\det(\mathbf{A})$ or $|\mathbf{A}|$ .

**Example 4.1.2**

Evaluate the determinant of the $3\times3$ matrix $\begin{pmatrix} 1 & 0 & -2 \\ 3 & 1 & 4 \\ 5 & 2 & -3 \end{pmatrix}$.

*Solution:*

Let **A** denote the given matrix. We evaluate the determinant by cofactor expansion along the first row:

$$\det(\mathbf{A}) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = 1\begin{vmatrix} 1 & 4 \\ 2 & -3 \end{vmatrix} + 0 + (-2)\begin{vmatrix} 3 & 1 \\ 5 & 2 \end{vmatrix} = -11 - 2 = -13.$$

**Example 4.1.3**

Evaluate the determinant of the matrix $\begin{pmatrix} 1 & 0 & -2 & 0 \\ 1 & 2 & 0 & 3 \\ 2 & 0 & 3 & 4 \\ 0 & -3 & 2 & 1 \end{pmatrix}$.

*Solution:*

Let **A** denote the given matrix. We evaluate the cofactor expansion of the first row:

$$\det(\mathbf{A}) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} + a_{14}C_{14} = 1\begin{vmatrix} 2 & 0 & 3 \\ 0 & 3 & 4 \\ -3 & 2 & 1 \end{vmatrix} + (-2)\begin{vmatrix} 1 & 2 & 3 \\ 2 & 0 & 4 \\ 0 & -3 & 1 \end{vmatrix}.$$

Now

$$\begin{vmatrix} 2 & 0 & 3 \\ 0 & 3 & 4 \\ -3 & 2 & 1 \end{vmatrix} = 2\begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix} + 3\begin{vmatrix} 0 & 3 \\ -3 & 2 \end{vmatrix} = -10 + 27 = 17,\ \begin{vmatrix} 1 & 2 & 3 \\ 2 & 0 & 4 \\ 0 & -3 & 1 \end{vmatrix} = -2\begin{vmatrix} 2 & 3 \\ -3 & 1 \end{vmatrix} - 4\begin{vmatrix} 1 & 2 \\ 0 & -3 \end{vmatrix} = -22 + 12 = -10.$$

So $\det(\mathbf{A}) = 17 + 20 = 37$.

**Definition**

A square matrix in which all the entries below (respectively above) the main diagonal are zeros is called an ***upper*** (respectively ***lower***) ***triangular matrix***. A square matrix in which all the entries off the main diagonal are zeros is called a ***diagonal matrix***.

For example, the matrices $\begin{pmatrix} 1 & -3 & 4 \\ 0 & 2 & 5 \\ 0 & 0 & 0 \end{pmatrix}$, $\begin{pmatrix} 8 & 0 & 0 \\ 1 & 2 & 0 \\ 3 & 3 & 4 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{pmatrix}$ are upper triangular matrix, lower triangular matrix and diagonal matrix respectively.

**Theorem 4.1.2**

If $\mathbf{A} = (a_{ij})$ is an $n \times n$ upper triangular, lower triangular or diagonal matrix, then $\det(\mathbf{A})$ is the product of the entries on the main diagonal of **A**, i.e. $\det(\mathbf{A}) = a_{11}a_{22}...a_{nn}$.

**A Special Rule to Find the Determinant of a $3 \times 3$ Matrix**

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32},$$

which can be memorised by the following mnemonic form

$$\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array}$$

where we sum the products of the entries on the right downward arrows then subtract the products of the entries on the left downward errors. (Does this look familiar?)

**Important: This only works for $3 \times 3$ matrix!**

## 4.2 Evaluating Determinants by Row Reduction

**Theorem 4.2.1**

If **A** is a square matrix with a row or a column of zeros, then $\det(\mathbf{A})=0$.

*Proof:*

Evaluating the determinant of **A** by cofactor expansion along that row or column of zeros, we can show $\det(\mathbf{A})=0$.

**Theorem 4.2.2**

If **A** is a square matrix, then $\det(\mathbf{A})=\det(\mathbf{A}^T)$.

You may prove this theorem by mathematical induction.

**Example 4.2.1**

Let $\mathbf{A}=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$, investigate the effects of the elementary row operations on its determinant.

*Investigation:*

(a) Multiplying a row by a scalar *k*

Consider the matrix $\mathbf{B}_1=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ ka_{21} & ka_{22} & ka_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Evaluating its determinant by cofactor expansion along the second row, we have

$$\det(\mathbf{B}_1)=(ka_{21})C_{21}+(ka_{22})C_{22}+(ka_{23})C_{23}=k(a_{21}C_{21}+a_{22}C_{22}+a_{23}C_{23})=k\det(\mathbf{A})$$

(b) Interchanging two rows

Consider the matrix $\mathbf{B}_2=\begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Evaluating its determinant by cofactor expansion along the third row, we have

$$\det(\mathbf{B}_2)=a_{31}\begin{vmatrix} a_{22} & a_{23} \\ a_{12} & a_{13} \end{vmatrix}-a_{32}\begin{vmatrix} a_{21} & a_{23} \\ a_{11} & a_{13} \end{vmatrix}+a_{33}\begin{vmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{vmatrix}=-a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}+a_{32}\begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix}-a_{33}\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$
$$=-(a_{31}C_{31}+a_{32}C_{32}+a_{33}C_{33})=-\det(\mathbf{A})$$

(c) Adding a multiple of one row to another row

Consider the matrix $\mathbf{B}_3 = \begin{pmatrix} a_{11}+ma_{31} & a_{12}+ma_{32} & a_{13}+ma_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$. Evaluating its determinant by cofactor expansion along the first row, we have

$$\begin{aligned}\det(\mathbf{B}_3) &= (a_{11}+ma_{31})C_{11} + (a_{12}+ma_{32})C_{12} + (a_{13}+ma_{33})C_{13} \\ &= (a_{11}C_{11}+a_{12}C_{12}+a_{13}C_{13}) + m(a_{31}C_{11}+a_{32}C_{12}+a_{33}C_{13}) \\ &= \det(\mathbf{A}) + \ldots = \ldots = \det(\mathbf{A})\end{aligned}$$

The following theorem describes the effect of an elementary row (or column) operation on the determinant of a matrix.

**Theorem 4.2.3**

Let **A** be a square matrix.

**(a)** If **B** is the matrix that results when a row (or a column) of **A** is multiplied by a scalar *k*, then $\det(\mathbf{B}) = k\det(\mathbf{A})$.

**(b)** If **B** is the matrix that results when two rows (or two columns) of **A** are interchanged, then $\det(\mathbf{B}) = -\det(\mathbf{A})$.

**(c)** If **B** is the matrix that results when a multiple of one row (or one column) of **A** is added to another row (or another column), then $\det(\mathbf{B}) = \det(\mathbf{A})$.

**Example 4.2.2**

Evaluate $\begin{vmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{vmatrix}$ using **Theorem 4.1.2** and **Theorem 4.2.3**.

*Solution:*

$$\begin{vmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{vmatrix} = -\begin{vmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{vmatrix} = -(1)(3)(-5) = 15.$$

- What can you say about the determinant of **A** if it has two identical rows (or columns)?

- What can you say about the determinant of *k***A**?

**Corollary 4.2.3.1**

If a square matrix **A** has two identical rows or two identical columns, then $\det(\mathbf{A}) = 0$.

**Corollary 4.2.3.2**

If **A** is a $n \times n$ square matrix, then for any scalar *k*, $\det(k\mathbf{A}) = k^n \det(\mathbf{A})$

### 4.3 Properties of Determinant

- Investigate whether each of the following statements is true given that **A** and **B** are square matrices of the same size:

  **(a)** $\det(\mathbf{A}+\mathbf{B}) = \det(\mathbf{A}) + \det(\mathbf{B})$.

  **(b)** $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$.

  **(c)** **A** is invertible if and only if $\det(\mathbf{A}) \neq 0$.

**Theorem 4.3.1**

Let **A**, **B** and **C** be $n \times n$ matrices that differs only in a single row, say the *r*th row, and suppose that the *r*th row of **C** is the sum of the corresponding entries in the *r*th rows of **A** and **B**. Then

$$\det(\mathbf{C}) = \det(\mathbf{A}) + \det(\mathbf{B}).$$

The same result hold for columns.

**Important:** $\det(\mathbf{A}+\mathbf{B}) \neq \det(\mathbf{A}) + \det(\mathbf{B})$ **in general!**

**Example 4.3.1**

Use 3 matrices to illustrate **Theorem 4.3.1**.

*Solution:*

**Theorem 4.3.2**

If **A** and **B** are $n \times n$ matrices, then

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

**Theorem 4.3.3**

A square matrix **A** is invertible if and only if $\det(\mathbf{A}) \neq 0$.

**Example 4.3.2**

Prove that if **A** is invertible, then $\det(\mathbf{A}^{-1}) = \dfrac{1}{\det(\mathbf{A})}$.

*Proof:*

By **Theorem 4.3.2**, $\det\det(\mathbf{A})\det(\mathbf{A}^{-1}) = \det(\mathbf{AA}^{-1}) = \det(\mathbf{I}) = 1$, by **Theorem 4.1.2**.

**Example 4.3.3**

Show that the matrix $\begin{pmatrix} 1 & -1 & 3 \\ 1 & 3 & 11 \\ -2 & 2 & -6 \end{pmatrix}$ is singular.

*Solution:*

Adding 2 times the first row to the third row, by **Theorem 4.2.3(c)**,

$\begin{vmatrix} 1 & -1 & 3 \\ 1 & 3 & 11 \\ -2 & 2 & -6 \end{vmatrix} = \begin{vmatrix} 1 & -1 & 3 \\ 1 & 3 & 11 \\ 0 & 0 & 0 \end{vmatrix} = 0$, by **Theorem 4.2.1**.

Hence the matrix is singular by **Theorem 4.3.3**.

## 4.4 Adjoint of a Matrix

**Definition**

Let $\mathbf{A} = \left(a_{ij}\right)$ be an $n \times n$ matrix and let $C_{ij}$ be the cofactor of $a_{ij}$. The matrix

$$\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix}$$

is called the ***matrix of cofactors*** from **A**.

The transpose of this matrix is called the adjoint of **A** and is denoted by $\text{adj}\left(\mathbf{A}\right)$.

**Example 4.4.1**

Find the adjoint of $\begin{pmatrix} 1 & 0 & -2 \\ 3 & 1 & 4 \\ 5 & 2 & -3 \end{pmatrix}$.

*Solution:*

The cofactors of the matrix are

$$\begin{matrix} C_{11} = -11 & C_{12} = 29 & C_{13} = 1 \\ C_{21} = -4 & C_{22} = 7 & C_{23} = -2 \\ C_{31} = 2 & C_{32} = -10 & C_{33} = 1 \end{matrix}$$

Therefore, the adjoint of the matrix is $\begin{pmatrix} -11 & -4 & 2 \\ 29 & 7 & -10 \\ 1 & -2 & 1 \end{pmatrix}$.

Using the adjoint of a matrix, we are not able to give a formula for the inverse of an invertible matrix, like the one for $2 \times 2$ matrix.

**Theorem 4.4.1**

If **A** is a square matrix, then

$$\mathbf{A}\text{adj}\left(\mathbf{A}\right) = \det\left(\mathbf{A}\right)\mathbf{I}.$$

In particular, if $\det\left(\mathbf{A}\right) \neq 0$, then **A** is invertible and

$$\mathbf{A}^{-1} = \frac{1}{\det\left(\mathbf{A}\right)}\text{adj}\left(\mathbf{A}\right).$$

*Proof:*

The sizes are consistent on both sides obviously. The $(i,\, j)$ entry of $\mathbf{A}\text{adj}(\mathbf{A})$ is

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \ldots a_{in}C_{jn}.$$

If $i = j$, this value is $\det(\mathbf{A})$.

If $i \neq j$, let $\mathbf{B}$ be the matrix by replacing the *j*th row in $\mathbf{A}$ with $\begin{pmatrix} a_{i1} & a_{i2} & \ldots & a_{in} \end{pmatrix}$.
By **Corollary 4.2.3.1**, we know $\det(\mathbf{B}) = 0$ as its *i*th and *j*th rows are identical.
The cofactors of the entries in the *j*th rows of $\mathbf{B}$ remains as $C_{j1}$, $C_{j2}$, …, $C_{jn}$ as they are computed when the *j*th row are deleted. Thus, $\det(\mathbf{B}) = a_{i1}C_{j1} + a_{i2}C_{j2} + \ldots a_{in}C_{jn}$ which is 0.

Therefore $\mathbf{A}\text{adj}(\mathbf{A})$ is a diagonal matrix whose entries on the main diagonals are all $\det(\mathbf{A})$. That is,

$$\mathbf{A}\text{adj}(\mathbf{A}) = \det(\mathbf{A})\mathbf{I}.$$

In particular, if $\det(\mathbf{A}) \neq 0$, multiplying both sides by $\mathbf{A}^{-1}$ on the left, we have, $\mathbf{A}^{-1}\left[\mathbf{A}\text{adj}(\mathbf{A})\right] = \mathbf{A}^{-1}\left[\det(\mathbf{A})\mathbf{I}\right] \Rightarrow \left(\mathbf{A}^{-1}\mathbf{A}\right)\text{adj}(\mathbf{A}) = \det(\mathbf{A})\left(\mathbf{A}^{-1}\mathbf{I}\right) \Rightarrow \mathbf{I}\text{adj}(\mathbf{A}) = \det(\mathbf{A})\mathbf{A}^{-1}$.
Thus $\mathbf{A}^{-1} = \dfrac{1}{\det(\mathbf{A})}\text{adj}(\mathbf{A})$.

**Example 4.4.2**

Use the result of **Example 4.4.1** and Theorem **4.4.1** to find the inverse of $\begin{pmatrix} 1 & 0 & -2 \\ 3 & 1 & 4 \\ 5 & 2 & -3 \end{pmatrix}$.

*Solution:*

$\det(\mathbf{A}) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = 1 \times (-11) + 0 + (-2) \times 1 = -13$.

So the inverse is $\mathbf{A}^{-1} = \dfrac{1}{\det(\mathbf{A})}\text{adj}(\mathbf{A}) = -\dfrac{1}{13}\begin{pmatrix} -11 & -4 & 2 \\ 29 & 7 & -10 \\ 1 & -2 & 1 \end{pmatrix}$.

## 4.5 Cramer's Rule

The following theorem gives a formula for the solution of some linear systems with $n$ equations and $n$ unknowns.

**Theorem 4.5.1 (Cramer's Rule)**

Let $\mathbf{Ax}=\mathbf{b}$ be a system of $n$ linear equations in $n$ unknowns such that $\det(\mathbf{A})\neq 0$. Then the linear system has exactly one solution, and the solution is given by

$$x_i=\frac{\det(\mathbf{A}_i)}{\det(\mathbf{A})},\ i=1,2,\ldots,n,$$

where $\mathbf{A}_i$ is the matrix obtained by replacing the $i$th column of $\mathbf{A}$ by $\mathbf{b}$.

**Example 4.5.1**

Solve the following system of linear equations using Cramer's Rule.

$$\begin{aligned} 3x &+5y & &=7 \\ 6x &+2y &+4z &=10 \\ -x &+4y &-3z &=0 \end{aligned}$$

*Solution:*

Evaluating the determinant of the coefficient matrix $\begin{pmatrix} 3 & 5 & 0 \\ 6 & 2 & 4 \\ -1 & 4 & -3 \end{pmatrix}$, we obtain $\det(\mathbf{A})=4$. The linear system has exactly one solution.

Now $\mathbf{A}_1=\begin{pmatrix} 7 & 5 & 0 \\ 10 & 2 & 4 \\ 0 & 4 & -3 \end{pmatrix}$, $\mathbf{A}_2=\begin{pmatrix} 3 & 7 & 0 \\ 6 & 10 & 4 \\ -1 & 0 & -3 \end{pmatrix}$ and $\mathbf{A}_3=\begin{pmatrix} 3 & 5 & 7 \\ 6 & 2 & 10 \\ -1 & 4 & 0 \end{pmatrix}$, their determinants are $-4$, 8 and 12 respectively. By Cramer's Rule, $x=\frac{-4}{4}=-1$, $y=\frac{8}{4}=2$ and $z=\frac{12}{4}=3$.

## §5 Real Vector Spaces

### 5.1 Definition of Real Vector Spaces

Consider $\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}$, we can think of elements in $\mathbb{R}^2$ algebraically as ordered pairs, or geometrically as 'vectors'. We can add any two elements in $\mathbb{R}^2$, and multiply any element in $\mathbb{R}^2$ by a scalar (real number), i.e.

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2) \text{ and } k(x, y) = (kx, ky) \text{, where } k \text{ is a real number.}$$

Similarly, for $\mathbf{M}_{2,2}(\mathbb{R})$, the set of all $2 \times 2$ matrices, we can add any two matrices and multiple a matrix by a scalar (real number), i.e.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a+e & b+f \\ c+g & d+h \end{pmatrix} \text{ and } k\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ka & kb \\ kc & kd \end{pmatrix} \text{, where } k \text{ is a real number.}$$

The two sets, $\mathbb{R}^2$ and $\mathbf{M}_{2,2}(\mathbb{R})$, together with addition and multiplication, share many common properties. In fact, there are many sets with addition and scalar multiplication defined on them that share these common properties. We shall make a general study of such system collectively.

**Definition**

A (***real***) ***vector space*** or (**real**) ***linear space*** is a nonempty set $V$ with two operations $\oplus$ and $\otimes$, called addition and (real) scalar multiplication, that satisfy all the following axioms:

**A1 (Closure under Addition):**

If **u** and **v** are in $V$, then $\mathbf{u} \oplus \mathbf{v} \in V$.

**A2 (Commutative Property for Addition):**

$$\mathbf{u} \oplus \mathbf{v} = \mathbf{v} \oplus \mathbf{u}.$$

**A3 (Associative Property for Addition):**

$$\mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w}) = (\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w}$$

**A4 (Additive Identity):**

There is an element **0** in $V$ such that $\mathbf{0} \oplus \mathbf{u} = \mathbf{u}$ and $\mathbf{u} \oplus \mathbf{0} = \mathbf{u}$ for all $\mathbf{u}$ in $V$. The element **0** is called the ***zero vector***.

**A5 (Additive Inverse):**

For each **u** in $V$, there exists an element $-\mathbf{u}$ in $V$, called the ***negative*** of **u**, such that $\mathbf{u} \oplus (-\mathbf{u}) = (-\mathbf{u}) \oplus \mathbf{u} = \mathbf{0}$.

For any real numbers $k$ and $l$,

**A6 (Closure under Scalar Multiplication)**

If $\mathbf{u}$ is in $V$, then $k \otimes \mathbf{u} \in V$.

**A7 (Distributive Property of Scalar Multiplication over Addition):**

$$k \otimes (\mathbf{u} \oplus \mathbf{v}) = k \otimes \mathbf{u} \oplus k \otimes \mathbf{v}.$$

**A8 (Distributive Property of Scalar Multiplication over Scalar Addition):**

$$(k + l) \otimes \mathbf{u} = k \otimes \mathbf{u} \oplus l \otimes \mathbf{u}.$$

**A9 (Associative Property for Multiplication):**

$$k \otimes (l \otimes \mathbf{u}) = (kl) \otimes \mathbf{u}.$$

**A10 (Multiplicative Identity):**

$$1 \otimes \mathbf{u} = \mathbf{u}.$$

If $V$ is a vector space, then the elements in $V$ are called ***vectors***.

**Important:**

The axioms of a vector space do not specify the nature of the vectors nor the operations.

Here are some examples of vector spaces.

**Example 5.1.1**

**(a)** $\mathbb{R}^2$, with the usual addition and scalar multiplication, is a vector space.
More generally, $\mathbb{R}^n$, with the usual addition and scalar multiplication, is a vector space.

**(b)** $\mathbf{M}_{2,2}(\mathbb{R})$, with the usual addition and scalar multiplication, is a vector space.
More generally, the set of all $m \times n$ real matrices $\mathbf{M}_{m,\mathrm{n}}(\mathbb{R})$ with the operations of matrix addition and scalar multiplication, is a vector space.

**(c)** Let $V$ be the set of all functions $\mathrm{f} : \mathbb{R} \to \mathbb{R}$. We define addition and scalar multiplication on $V$ as follows: For $\mathrm{f}, \mathrm{g} \in V$ and $k \in \mathbb{R}$, $(\mathrm{f} + \mathrm{g})(x) = \mathrm{f}(x) + \mathrm{g}(x)$, $(k\mathrm{f})(x) = k\mathrm{f}(x)$.

**Example 5.1.2**

Let $\mathbf{P}_2$ denote the set of all polynomials with real coefficients of degree less or equal to 2, i.e.

$$\mathbf{P}_2 = \left\{a + bx + cx^2 : a, b, c \in \mathbb{R}\right\}.$$

Show that, $\mathbf{P}_2$ with the usual addition and scalar multiplication of polynomials, is a vector space.

*Proof:*

We need to verify that it satisfies the ten axioms.
Let $\mathbf{u} = a + bx + cx^2 \in \mathbf{P}_2$, $\mathbf{v} = d + ex + fx^2 \in \mathbf{P}_2$ and $\mathbf{w} = g + hx + ix^2 \in \mathbf{P}_2$, and $k, l \in \mathbb{R}$.

**A1** $\mathbf{u} + \mathbf{v} = \left(a + bx + cx^2\right) + \left(d + ex + fx^2\right) = (a + d) + (b + e)x + (c + f)x^2 \in \mathbf{P}_2$.

**A2** $\mathbf{v} + \mathbf{u} = \left(d + ex + fx^2\right) + \left(a + bx + cx^2\right) = \left(a + bx + cx^2\right) + \left(d + ex + fx^2\right) = \mathbf{u} + \mathbf{v}$.

**A3** $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = \left(a + bx + cx^2\right) + \left[\left(d + ex + fx^2\right) + \left(g + hx + ix^2\right)\right]$.

$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \left[\left(a + bx + cx^2\right) + \left(d + ex + fx^2\right)\right] + \left(g + hx + ix^2\right)$.

Thus $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (a + d + g) + (b + e + h)x + (c + f + i)x^2 = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$.

**A4** Let $\mathbf{0} = 0 + 0x + 0x^2 \in \mathbf{P}_2$ then $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = a + bx + cx^2 = \mathbf{u}$.

**A5** Let $-\mathbf{u} = -a - bx - cx^2 \in \mathbf{P}_2$, then $\mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = 0 + 0x + 0x^2 = \mathbf{0}$.

**A6** $k\mathbf{u} = k\left(a + bx + cx^2\right) = ka + (kb)x + (kc)x^2 \in \mathbf{P}_2$.

**A7** $k(\mathbf{u} + \mathbf{v}) = k\left[\left(a + bx + cx^2\right) + \left(d + ex + fx^2\right)\right] = k\left(a + bx + cx^2\right) + k\left(d + ex + fx^2\right) = k\mathbf{u} + k\mathbf{v}$.

**A8** $(k + l)\mathbf{u} = (k + l)\left(a + bx + cx^2\right) = k\left(a + bx + cx^2\right) + l\left(a + bx + cx^2\right) = k\mathbf{u} + l\mathbf{u}$.

**A9** $k(l\mathbf{u}) = k\left[l\left(a + bx + cx^2\right)\right] = (kl)\left(a + bx + cx^2\right) = (kl)\mathbf{u}$.

**A10** $1\mathbf{u} = 1\left(a + bx + cx^2\right) = a + bx + cx^2 = \mathbf{u}$.

Therefore, $\mathbf{P}_2$ with the usual addition and scalar multiplication of polynomials, is a vector space.

More generally, let $\mathbf{P}_n$ be the set of all polynomials with real coefficients of degree less or equal to $n$. Then $\mathbf{P}_n$ with the usual addition and scalar multiplication of polynomials, is a vector space.

- Is the set of all polynomials with real coefficient a vector space under the usual addition and multiplication of polynomials?

> **Definition**
>
> A ***trivial vector space*** or ***zero vector space*** contains only the zero vector, i.e. $\{\mathbf{0}\}$ with the addition $\oplus$ and scalar multiplication $\otimes$ defined by
>
> $$\mathbf{0} \oplus \mathbf{0} = \mathbf{0} \text{ and } k \otimes \mathbf{0} = \mathbf{0}.$$

- Explain why a trivial vector space is a vector space.

**Example 5.1.3**

Let $V=\mathbb{R}^2$, and define addition $\oplus$ and scalar multiplication $\otimes$ on $V$ by

$$(a,b)\oplus(c,d)=(a+c,b+d+1),\ k\otimes(a,b)=(ka,kb+k-1).$$

Show that $V$ is a vector space under $\oplus$ and $\otimes$.

*Proof:*

We need to verify that it satisfies the ten axioms.
Let $\mathbf{u}=(a,b)\in V$, $\mathbf{v}=(c,d)\in V$ and $\mathbf{w}=(e,f)\in V$, and $k,l\in\mathbb{R}$.

**A1** $\mathbf{u}\oplus\mathbf{v}=(a,b)\oplus(c,d)=(a+c,b+d+1)\in V$.

**A2** $\mathbf{v}\oplus\mathbf{u}=(c,d)\oplus(a,b)=(c+a,d+b+1)=(a+c,b+d+1)=\mathbf{u}\oplus\mathbf{v}$.

**A3** $\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})=(a,b)\oplus\left[(c,d)\oplus(e,f)\right]=(a,b)\oplus(c+e,d+f+1)=(a+c+e,b+d+f+2)$.
$(\mathbf{u}\oplus\mathbf{v})\oplus\mathbf{w}=(a+c,b+d+1)\oplus(e,f)=(a+c+e,b+d+f+2)$.
Thus $\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})=(\mathbf{u}\oplus\mathbf{v})\oplus\mathbf{w}$.

**A4** Let $\mathbf{0}=(0,-1)\in V$. Then $\mathbf{u}\oplus\mathbf{0}=(a,b)\oplus(0,-1)=(a+0,b-1+1)=(a,b)=\mathbf{u}\oplus\mathbf{0}=\mathbf{u}$.

**A5** Let $-\mathbf{u}=(-a,-b-2)\in V$, then
$\mathbf{u}\oplus(-\mathbf{u})=(-\mathbf{u})\oplus\mathbf{u}=(a,b)\oplus(-a,-b-2)=(a-a,b-b-2+1)=(0,-1)=\mathbf{0}$.

**A6** $k\otimes\mathbf{u}=k\otimes(a,b)=(ka,kb+k-1)\in V$.

**A7** $k\otimes(\mathbf{u}\oplus\mathbf{v})=k\otimes(a+c,b+d+1)=(k(a+c),k(b+d+1)+k-1)=(ka+kc,kb+kd+2k-1)$.
$k\otimes\mathbf{u}\oplus k\otimes\mathbf{v}=(ka,kb+k-1)\oplus(kc,kd+k-1)=(ka+kc,kb+kd+k-1+k-1+1)$.
Thus $k\otimes(\mathbf{u}\oplus\mathbf{v})=k\otimes\mathbf{u}\oplus k\otimes\mathbf{v}$.

**A8** $(k+l)\otimes\mathbf{u}=((k+l)a,(k+l)b+(k+l)-1)=(ka+la,kb+lb+k+l-1)$.
$k\otimes\mathbf{u}\oplus l\otimes\mathbf{u}=(ka,kb+k-1)\oplus(la,lb+l-1)=(ka+la,kb+k-1+lb+l-1+1)$.
Thus, $(k+l)\otimes\mathbf{u}=k\otimes\mathbf{u}\oplus l\otimes\mathbf{u}$.

**A9** $k\otimes(l\otimes\mathbf{u})=k\otimes(la,lb+l-1)=(kla,k(lb+l-1)+k-1)=(kla,klb+kl-1)$.
$(kl)\otimes\mathbf{u}=(kla,klb+kl-1)$.
Thus, $k\otimes(l\otimes\mathbf{u})=(kl)\otimes\mathbf{u}$

**A10** $1\otimes\mathbf{u}=1\otimes(a,b)=(a,b+1-1)=(a,b)=\mathbf{u}$.

Therefore $V$ is a vector space under $\oplus$ and $\otimes$.

**Example 5.1.4**

Let $U=\{(x,y):xy=0\}$. Show that $U$ is not a vector space under usual addition and scalar multiplication.

*Proof:*

We just need to identify an axiom that it fails to satisfy.

**A1** Let $\mathbf{u}=(1,0)$ and $\mathbf{v}=(0,1)$, we see that $\mathbf{u},\mathbf{v}\in U$ but $\mathbf{u}+\mathbf{v}=(1,1)\notin U$, so **A1** fails.

- Can you figure out another axiom that *U* under usual addition and scalar multiplication fails to satisfy?

**Example 5.1.5**

Determine whether each of the following is a vector space.

**(a)** *W* under the usual matrix addition and scalar multiplication, where
$$W=\left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix}: a+b\le c+d\right\}.$$

**(b)** *Q* under usual addition and scalar multiplication, where
$$Q=\{(x,y,z): x+2y-3z=0\}.$$

**(c)** *F* under usual addition and scalar multiplication, where
$$F=\{\mathrm{f}:\mathbb{R}\to\mathbb{R}:\mathrm{f}'+\mathrm{f}=0\}.$$

**(d)** $S=\{-1,0,1\}$**, and define addition $\oplus$ and scalar multiplication $\otimes$ on** *S* by

For $a,b\in S$ and $k\in\mathbb{R}$, $a\oplus b=ab$, $k\otimes a=\begin{cases}0 & \text{if } k=0\\ a & \text{if } k\neq 0\end{cases}$.

*Solution:*

**(b)** and **(c)** are vector spaces.

**(a)** is not a vector space as **A5** fails. For example, let $\mathbf{w}=\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}\in W$.

We cannot find $-\mathbf{w}$ in *W*, as $\mathbf{0}$ has to be $\begin{pmatrix}0 & 0\\ 0 & 0\end{pmatrix}$ but the only $2\times 2$ matrix that satisfies the property of $-\mathbf{w}$ is $\begin{pmatrix}-1 & -2\\ -3 & -4\end{pmatrix}$ which is not in *W* ($(-1)+(-2)>(-3)+(-4)$.

**(d)** is not a vector space as **A8** fails. For example, let $a=1$, $k=2$ and $l=-2$.
$(k\oplus l)\otimes\mathbf{u}=[2+(-2)]\otimes 1=0\otimes 1=0$, but $k\otimes\mathbf{u}\oplus l\otimes\mathbf{u}=2\otimes 1\oplus(-2)\otimes 1=1\oplus 1=1$.

- Can you figure out other axioms that the non-examples do not satisfy?

## 5.2 Basic Properties of Vector Spaces

We shall now state and prove some basic properties of vector space. Note that the proofs of these properties use only the axioms of vector spaces, and NOT specific properties of any concrete vector space such as $\mathbb{R}^2$ (thus we cannot assume $V = \mathbb{R}^2$ or let $\mathbf{v} = (a, b)$ in our proofs).

**Lemma 5.2.1**

Let $V$ be a vector space and let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors in $V$. If $\mathbf{u} \oplus \mathbf{v} = \mathbf{u} \oplus \mathbf{w}$ or $\mathbf{v} \oplus \mathbf{u} = \mathbf{w} \oplus \mathbf{u}$, then $\mathbf{v} = \mathbf{w}$.

This lemma allows us to 'subtract' the same vectors from both sides of an identity.

*Proof:*

Suppose that $\mathbf{u} \oplus \mathbf{v} = \mathbf{u} \oplus \mathbf{w}$. By **A5**, $\mathbf{u}$ has a negative, $-\mathbf{u} \in V$. Adding $-\mathbf{u}$ to both sides on the left, we have

$$-\mathbf{u} \oplus (\mathbf{u} \oplus \mathbf{v}) = -\mathbf{u} \oplus (\mathbf{u} \oplus \mathbf{w})$$

$$(-\mathbf{u} \oplus \mathbf{u}) \oplus \mathbf{v} = (-\mathbf{u} \oplus \mathbf{u}) \oplus \mathbf{w} \text{ (by } \mathbf{A3}\text{)}$$

$$\mathbf{0} \oplus \mathbf{v} = \mathbf{0} \oplus \mathbf{w} \text{ (by } \mathbf{A5}\text{)}$$

$$\mathbf{v} = \mathbf{w} \text{ (by } \mathbf{A4}\text{)}$$

The proof of the case when $\mathbf{v} \oplus \mathbf{u} = \mathbf{w} \oplus \mathbf{u}$ is similar.

**Theorem 5.2.2**

**(a)** (Uniqueness of $\mathbf{0}$) The vector $\mathbf{0} \in V$ is the unique additive identity for any $\mathbf{u} \in V$.
**(b)** (Uniqueness of $-\mathbf{u}$) The vector $-\mathbf{u} \in V$ is the unique additive inverse for a given $\mathbf{u} \in V$.
**(c)** $0 \otimes \mathbf{u} = \mathbf{0}$ for any $\mathbf{u} \in V$.
**(d)** $k \otimes \mathbf{0} = \mathbf{0}$ for any $k \in \mathbb{R}$.
**(e)** $(-1) \otimes \mathbf{u} = -\mathbf{u}$ for any $\mathbf{u} \in V$.
**(f)** If $k \otimes \mathbf{u} = \mathbf{0}$, then either $k = 0$ or $\mathbf{u} = \mathbf{0}$.

*Proof:*

**(a)** Suppose $\mathbf{w} \oplus \mathbf{u} = \mathbf{u}$.

$$\mathbf{w} \oplus \mathbf{u} = \mathbf{0} \oplus \mathbf{u} \text{ (by } \mathbf{A4}\text{)}$$

$$\mathbf{w} = \mathbf{0} \text{ (by } \mathbf{Lemma\ 5.2.1}\text{)}$$

**(b)** Suppose $\mathbf{u} \oplus \mathbf{w} = \mathbf{0}$.

$$\mathbf{u} \oplus \mathbf{w} = \mathbf{u} \oplus (-\mathbf{u}) \text{ (by } \mathbf{A5}\text{)}$$

$$\mathbf{w} = -\mathbf{u} \text{ (by } \mathbf{Lemma\ 5.2.1}\text{)}$$

**(c)**

$$0 \otimes \mathbf{u} = 0 \otimes \mathbf{u}$$

$$0 \otimes \mathbf{u} \oplus \mathbf{0} = (0 + 0) \otimes \mathbf{u} \text{ (by } \mathbf{A4}\text{)}$$

$$0\otimes\mathbf{u}\oplus\mathbf{0}=0\otimes\mathbf{u}\oplus 0\otimes\mathbf{u} \text{ (by } \mathbf{A8}\text{)}$$
$$\mathbf{0}=0\otimes\mathbf{u} \text{ (by } \mathbf{Lemma\ 5.2.1}\text{)}$$

**(d)**

$$k\otimes\mathbf{u}=k\otimes\mathbf{u}$$
$$k\otimes(\mathbf{u}\oplus\mathbf{0})=k\otimes\mathbf{u}\oplus\mathbf{0} \text{ (by } \mathbf{A4}\text{)}$$
$$k\otimes\mathbf{u}\oplus k\otimes\mathbf{0}=k\otimes\mathbf{u}\oplus\mathbf{0} \text{ (by } \mathbf{A7}\text{)}$$
$$k\otimes\mathbf{0}=\mathbf{0} \text{ (by } \mathbf{Lemma\ 5.2.1}\text{)}$$

**(e)**

$$\mathbf{u}\oplus(-\mathbf{u})=\mathbf{0} \text{ (by } \mathbf{A5}\text{)}$$
$$\mathbf{u}\oplus(-\mathbf{u})=0\otimes\mathbf{u} \text{ (by } \mathbf{Theorem\ 5.2.2(c)}\text{)}$$
$$\mathbf{u}\oplus(-\mathbf{u})=\left[1+(-1)\right]\otimes\mathbf{u}$$
$$\mathbf{u}\oplus(-\mathbf{u})=1\otimes\mathbf{u}\oplus(-1)\otimes\mathbf{u} \text{ (by } \mathbf{A8}\text{)}$$
$$\mathbf{u}\oplus(-\mathbf{u})=\mathbf{u}\oplus(-1)\otimes\mathbf{u} \text{ (by } \mathbf{A10}\text{)}$$
$$(-\mathbf{u})=(-1)\otimes\mathbf{u} \text{ (by } \mathbf{Lemma\ 5.2.1}\text{)}$$

**(f)** Given that $k\otimes\mathbf{u}=\mathbf{0}$, if $k\neq 0$,

$$\frac{1}{k}\otimes(k\otimes\mathbf{u})=\left(k\times\frac{1}{k}\right)\otimes\mathbf{u} \text{ (by } \mathbf{A9}\text{)}$$
$$\frac{1}{k}\otimes\mathbf{0}=1\otimes\mathbf{u} \text{ (given)}$$
$$\mathbf{0}=\mathbf{u} \text{ (by } \mathbf{Theorem\ 5.2.2(d)} \text{ and } \mathbf{A10}\text{)}$$

Therefore, if $k\otimes\mathbf{u}=\mathbf{0}$, then $k=0$ or $\mathbf{u}=\mathbf{0}$.

## 5.3 Subspaces

Consider the set $U=\{(x,0):x\in\mathbb{R}\}$. It can be easily verified that *U* is a vector space under the usual addition and scalar multiplication. Note that *U* is a subset of $\mathbb{R}^2$, and $\mathbb{R}^2$ is a vector space under the same operations as that on *U*. We shall now defined a term to describe in general a relation between two vector spaces such as that between *U* and $\mathbb{R}^2$.

> **Definition**
>
> Let *V* be a vector space and let *W* be a <u>nonempty</u> subset of *V*. Then *W* is called a subspace of *V* if *W* itself is a vector space under the <u>same</u> addition and scalar multiplication defined on *V*.

For example, the subset *U* of $\mathbb{R}^2$ above is a subspace of $\mathbb{R}^2$.

Given a nonempty subset *W* of a vector space *V*, to prove *W* is a subspace of *V*, by right we should show that *W* satisfies all the ten vector space axioms under the addition and scalar multiplication defined on *V*, which is tedious.

- Which axiom(s) do we need to verify for *W*, knowing *W* is a subset of a vector space *V*?

**Theorem 5.3.1**

Let $W$ be a nonempty subset of a vector space $V$. Then $W$ is a subspace of $V$ if and only if it satisfies both of the following conditions:

**(a)** For all $\mathbf{u}$ and $\mathbf{v}$ in $W$, $\mathbf{u} \oplus \mathbf{v}$ is in $W$ (we say that $W$ is closed under addition).

**(b)** For all $\mathbf{u}$ in $W$ and all scalars $k$, $k \otimes \mathbf{u}$ is in $W$ (we say that $W$ is closed under scalar multiplication.

The 'only if' part is obviously true because of the definition of vector space.

To show the 'if' part, we need to show that the other 8 axioms are definitely true when **(a)** and **(b)** hold for the nonempty subset $W$.

**Example 5.3.1**

It is given that $V$ is a vector space under $\oplus$ and $\otimes$. Show that for any $W \subseteq V$ and $W \neq \Phi$,

$\mathbf{0} \in W$ if $W$ is closed under addition and scalar multiplication.

*Proof:*

Since $W \neq \varnothing$, we can find an element $\mathbf{u} \in W$.
Since $W \subseteq V$, $\mathbf{u} \in V$. By **Theorem 5.2.2(c)**, $0 \otimes \mathbf{u} = \mathbf{0}$.
Since $W$ is closed under scalar multiplication, $0 \otimes \mathbf{u} \in W$.
Therefore $\mathbf{0} \in W$.

- For any vector space $V$, $\{\mathbf{0}\}$ and $V$ are its subspaces.

**Example 5.3.2**

Let $W = \{(x, y) : 2x - y = 0\}$. Show that $W$ is a subspace of $\mathbb{R}^2$ under the same usual addition and scalar multiplication.

*Proof:*

Since $2(0) - 0 = 0$, $(0,0) \in W$. Thus $W \neq \Phi$.

Let $\mathbf{u} = (a,b)$ and $\mathbf{v} = (c,d)$ be any elements in $W$. We know $2a - b = 0$ and $2c - d = 0$.

$\mathbf{u} + \mathbf{v} = (a+c, b+d)$. Since $2(a+c) - (b+d) = (2a-b) + (2c-d) = 0$, we know $\mathbf{u} + \mathbf{v} \in W$.

$k\mathbf{u} = (ka, kb)$. Since $2(ka) - (kb) = k(2a-b) = 0$, we know $k\mathbf{u} \in W$.

By **Theorem 5.3.1**, $W$ is a subspace of $\mathbb{R}^2$.

- What is the geometrical interpretation of **Example 5.3.2**?
- Is $W' = \{(x, y) : 2x - y = 1\}$ also a subspace of $\mathbb{R}^2$?

**Important:**

Now we can shorten the proof to show that $W$ is a vector space sometimes:

Step 1: Explain that $W$ is a subset of a well-known vector space.
Step 2: Show that $W$ is nonempty (by finding an element in $W$, usually $\mathbf{0}$).
Step 3: Show that $W$ is closed under both addition and scalar multiplication.

But note this method will not work if it is not obvious that $W$ is a subset of a well-known vector space.

**Example 5.3.3**

Show that $W=\{(x,y):x\geq 0\}$ is not a subspace of $\mathbb{R}^2$.

*Proof:*

$W$ is not closed under scalar multiplication. For example, $(1,0)\in W$ but $(-1)(1,0)=(-1,0)\notin W$. Therefore $W$ is not subspace of $\mathbb{R}^2$ by **Theorem 5.3.1**.

**Example 5.3.4**

Explain whether $U=\{\mathbf{A}\in\mathbf{M}_{2,2}(\mathbb{R}):\mathbf{A}^T=\mathbf{A}\}$ is a subspace of $\mathbf{M}_{2,2}(\mathbb{R})$.

*Solution:*

Yes it is.

Since $\mathbf{O}_{2\times 2}=(\mathbf{O}_{2\times 2})^T$, $\mathbf{O}_{2\times 2}\in U$ so $U$ is nonempty

Let $\mathbf{A}$, $\mathbf{B}$ be any two elements in $U$, and $k\in\mathbb{R}$.

By **Theorem 2.3.4**,

$$(\mathbf{A}+\mathbf{B})^T=\mathbf{A}^T+\mathbf{B}^T=\mathbf{A}+\mathbf{B} \text{ and } (k\mathbf{A})^T=k\left(\mathbf{A}^T\right)=k\mathbf{A}.$$

Thus both $\mathbf{A}+\mathbf{B}$ and $k\mathbf{A}$ are in $U$.

Therefore $U$ is a subspace of $\mathbf{M}_{2,2}(\mathbb{R})$ by **Theorem 5.3.1**.

## §6 Span, Linear Independence, Basis and Dimension

### 6.1 Span

Let $W=\{(a,b,c):a,b,c\in\mathbb{R},a+b=c\}$. Then it can be verified easily that *W* is a subspace of $\mathbb{R}^3$. Note that *W* is an infinite set. Is there some way to represent the vectors in *W* using a *finite number of fixed vectors in W*?

Let $\mathbf{v}_1=(1,0,1)$ and $\mathbf{v}_2=(0,1,1)$ be two vectors in *W*.

Now consider another vector $(1,1,2)$ in *W*, we can write

$$(1,1,2)=1(1,0,1)+1(0,1,1)=1\mathbf{v_1}+1\mathbf{v_2}.$$

**Example 6.1.1**

Show that any vector $\mathbf{u}\in W$ can be written in the form $\alpha\mathbf{v}_1+\beta\mathbf{v}_2$ for some $\alpha,\beta\in\mathbb{R}$.

*Proof:*

$\mathbf{u}=(a,b,c)=(a,b,a+b)=a(1,0,1)+b(0,1,1)=\alpha\mathbf{v}_1+\beta\mathbf{v}_2$, where $\alpha=a$ and $\beta=b$.

Thus we can use two fixed vectors $\mathbf{v}_1=(1,0,1)$ and $\mathbf{v}_2=(0,1,1)$ in *W* to represent an arbitrary vector in *W*.

In general, given a vector space *V*, is it possible to represent *V* using a finite number of fixed vectors in *V*, in the sense of the example above? To facilitate the discussion of this, we need to introduce some technical terms.

**Definition**

Let *V* be a vector space and let $\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n$ be vectors in *V*. A vector $\mathbf{v}$ in *V* is called a ***linear combination*** of the vectors of the vectors $\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_n$ if there are scalars $k_1,k_2,...,k_n$ such that

$$\mathbf{v}=k_1\otimes\mathbf{v}_1\oplus k_2\otimes\mathbf{v}_2\oplus...\oplus k_n\otimes\mathbf{v}_n\ .$$

**Example 6.1.2**

In $\mathbb{R}^3$, determine whether each of the vectors

**(a)** $(1,0,9)$ and

**(b)** $(1,5,1)$,

is a linear combination of $(1,2,-1)$ and $(3,5,2)$.

*Solution:*

**(a)** To determine whether $(1,0,9)$ is a linear combination of $(1,2,-1)$ and $(3,5,2)$, we need to check whether the vector equation

$$(1,0,9)=k(1,2,-1)+l(3,5,2)$$

has a solution in *k* and *l*. This equation gives us a system of linear equations:

$$\begin{aligned} k &\quad +3l &= 5 \\ 2k &\quad +5l &= 7 \\ -k &\quad +2l &= 3 \end{aligned}$$

Solving the linear system (for example, by Gaussian elimination), we obtain the solution $k=-5$, $l=2$. Since $(1,0,9)=(-5)(1,2,-1)+(2)(3,5,2)$, we conclude that $(1,0,9)$ is a linear combination of $(1,2,-1)$ and $(3,5,2)$.

**(b)** Similarly, consider the vector equation

$$(1,5,1)=m(1,2,-1)+n(3,5,2).$$

This leads to the linear system:

$$\begin{aligned} m &\quad +3n &= 1 \\ 2m &\quad +5n &= 5 \\ -m &\quad +2l &= 1 \end{aligned}$$

The augmented matrix of the linear system is $\begin{pmatrix} 1 & 3 & 1 \\ 2 & 5 & 5 \\ -1 & 2 & 1 \end{pmatrix}$.

Performing elementary row operations on this matrix gives its row-echelon form $\begin{pmatrix} 1 & 3 & 1 \\ 0 & 1 & -3 \\ 0 & 0 & 17 \end{pmatrix}$.

It is clear that the linear system has no solution as the last equation now is $0m+0n=17$.
Hence $(1,5,1)$ is not a linear combination of $(1,2,-1)$ and $(3,5,2)$.

In the above example, we ask whether a particular vector is a linear combination of a set of vectors. Now we want to study whether every vector in a vector space is a linear combination of a set of vectors.

**Definition**

Let $V$ be a vector space and let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ be vectors in *V*. We say that $V$ is ***spanned*** by $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ (or $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ ***span*** *V*, or $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ is a ***spanning set*** for *V* equivalently) if every vector in *V* is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$.

If *V* is spanned by $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$, then we write $V=\text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$.

If *V* is spanned by $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ in *V*, then

$$V=\text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}=\{k_1\otimes\mathbf{v}_1\oplus k_2\otimes\mathbf{v}_2\oplus ...\oplus k_n\otimes\mathbf{v}_n : k_1, k_2, ..., k_n\in\mathbb{R}\}.$$

**Example 6.1.3**

In **Example 6.1.1**, we can say $W=\{(a,b,c): a,b,c\in\mathbb{R}, a+b=c\}$ has a spanning set $\{(1,0,1),(0,1,1)\}$. Find another spanning set for *W*.

*Solution:*

Since $(a,b,c)=(a,c-a,c)=a(1,-1,0)+c(0,1,1)$, *W* has another spanning set $\{(1,-1,0),(0,1,1)\}$.

- The spanning set for a vector space *need not be unique.*
- Can we say that $\{(1,-1,0),(0,1,1),(1,0,1)\}$ is another spanning set for *W*?
- Can we say that $\{(1,-1,0),(0,1,1),(1,0,0)\}$ is another spanning set for *W*?

**Example 6.1.4**

Determine whether $\mathbf{P}_2$ is spanned by the vectors $1+x-2x^2$, $-3x+x^2$.

*Solution:*

No. Consider $1+x\in\mathbf{P}_2$.

Let $1+x=k\left(1+x-2x^2\right)+l\left(-3x+x^2\right)$. Comparing the coefficients, we have

$$\begin{aligned} k&=1 && \text{...(1)}\\ k-3l&=1 && \text{...(2)}\\ -2k+l&=0 && \text{...(3)} \end{aligned}$$

Substituting (1) into (2), we have $l=0$, but $k=1$ and $l=0$ do not satisfy (3), so there is no solution to the linear system.

Since $1+x\in\mathbf{P}_2$ cannot be written as a linear combination of the two given vectors, $\mathbf{P}_2$ is not spanned by the two given vectors.

**Example 6.1.5**

Find a spanning set for the subspace $V=\{(a,b,c,d):\ a+b-c=0, a+2c-d=0\}$ of $\mathbb{R}^4$.

*Solution:*

Take an arbitrary vector $(a,b,c,d)$ in *V*, so

$$\begin{array}{lllll} a & +b & -c & & =0\\ a & & +2c & -d & =0 \end{array}$$

By Gauss-Jordan elimination, the above linear system is equivalent to

$$\begin{array}{llll} a & & +2c & -d & =0 \\ & b & -3c & +d & =0 \end{array}$$

Let $c=s$ and $d=t$, where $s$ and $t$ are real numbers. Then we obtain the general solution of the linear system: $a=-2s+t$, $b=3s-t$, $c=s$ and $d=t$. Thus,

$$(a,b,c,d)=(-2s+t,3s-t,s,t)=s(-2,3,1,0)+t(1,-1,0,1).$$

So every vector in $V$ is a linear combination of $(-2,3,1,0)$ and $(1,-1,0,1)$. As these vectors lie in $V$. We conclude that $\{(-2,3,1,0),(1,-1,0,1)\}$ is a spanning set for $V$.

Consider the vector $(1,2,1)$ in $\mathbb{R}^3$. Can we find a subspace of $\mathbb{R}^3$ containing $(1,2,1)$ that is as "small" as possible?

**Theorem 6.1.1**

Let $V$ be a vector space can let $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ be vectors in $V$. Let $W$ be the subset of $V$ defined by

$$W=\{k_1 \otimes \mathbf{v}_1 \oplus k_2 \otimes \mathbf{v}_2 \oplus ... \oplus k_n \otimes \mathbf{v}_n : k_1, k_2, ..., k_n \in \mathbb{R}\}.$$

Then $W$ is a subspace of $V$ containing $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$. Furthermore, $W$ is the "smallest" subspace that contains these vectors, in the sense that if $U$ is a subspace of $V$ and $U$ also contains $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$, then $W \subseteq U$.

Note that $W=\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$.

**Example 6.1.6**

Let $1+x$, $x-x^2$ be vectors in $\mathbf{P}_2$. State the smallest subspace of $\mathbf{P}_2$ that contains these two vectors.

*Solution:*

$\{a(1+x)+b(x-x^2) : a,b \in \mathbb{R}\}$.

- Let $\mathbf{u}$ and $\mathbf{v}$ be nonzero vectors in $\mathbb{R}^3$, where $\mathbf{u}$ is not a scalar multiple of $\mathbf{v}$. Geometrically, what is $\operatorname{span}\{\mathbf{u}\}$ and what is $\operatorname{span}\{\mathbf{u},\mathbf{v}\}$?

## 6.2 Linear Independence

Consider **Example 6.1.3**, we have obtained $\{(1,-1,0),(0,1,1)\}$ as a spanning set for *W*. Let *S* denote this spanning set.

We may also say that $\{(1,-1,0),(0,1,1),(1,0,1)\}$ is another spanning set of *W*. Let *T* denote this spanning set.

**S** is a "smaller" spanning set than *T* in the sense that it has fewer vectors in *T*. Note that *S* is obtained from *T* by deleting the vector $(1,0,1)$.

- Can we delete any vector from **S** to get an even 'smaller' spanning set for *W*?

In general, given a spanning set *S* for a vector space *V*, can we reduce a number of vectors in *S* to get a "smaller" spanning set for *V*? To help answer this question, we introduce the following concept.

> **Definition**
>
> A set of vectors $S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n\}$, $r>1$, is called ***linearly dependent***, if one of the vector in *S* is a linear combination of the <u>other</u> vectors in *S*, otherwise it is called ***linearly independent***, i.e. <u>none</u> of the vectors in *S* is a linear combination of the other vectors in *S*.
>
> If $S=\{\mathbf{v}\}$, then *S* is linearly independent if $\mathbf{v}\neq\mathbf{0}$, and linearly dependent if $\mathbf{v}=\mathbf{0}$.

For example, $T=\{(1,-1,0),(0,1,1),(1,0,1)\}$ is linearly dependent as

$$(1,-1,0)=(-1)(0,1,1)+(1)(1,0,1).$$

$S=\{(1,-1,0),(0,1,1)\}$ is linearly independent as neither vector is a multiple of the other.

**Example 6.2.1**

Let $V$ be a vector space and suppose $S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}$ is a set of vectors in *V*. It is given that $V=\operatorname{span}(S)$. Show that

**(a)** if $\mathbf{v}_1$ is a linear combination of $\{\mathbf{v}_2,\ldots,\mathbf{v}_k\}$, then $V=\operatorname{span}\{\mathbf{v}_2,\ldots,\mathbf{v}_k\}$;

**(b)** if **S** is linearly independent and *T* is a set obtained by removing one vector from *S*, prove that *T* does not span *V*.

*Solution:*

**(a)** Let $\mathbf{v}_1 = \alpha_2 \otimes \mathbf{v}_2 \oplus \alpha_3 \otimes \mathbf{v}_3 \oplus \ldots \oplus \alpha_k \otimes \mathbf{v}_k$, where $\alpha_2, \alpha_3, \ldots, \alpha_k \in \mathbb{R}$.

Take any vector **u** in *V*, since *S* spans *V*, **u** must be a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, i.e.

$$\begin{aligned}\mathbf{u} &= \beta_1 \otimes \mathbf{v}_1 \oplus \beta_2 \otimes \mathbf{v}_2 \oplus \ldots \oplus \beta_k \otimes \mathbf{v}_k \\ &= \beta_1 \otimes \left(\alpha_2 \otimes \mathbf{v}_2 \oplus \alpha_3 \otimes \mathbf{v}_3 \oplus \ldots \oplus \alpha_k \otimes \mathbf{v}_k\right) \oplus \beta_2 \otimes \mathbf{v}_2 \oplus \ldots \oplus \beta_k \otimes \mathbf{v}_k \\ &= \left(\alpha_2\beta_1 + \beta_2\right) \otimes \mathbf{v}_2 \oplus \left(\alpha_3\beta_1 + \beta_3\right) \otimes \mathbf{v}_3 \oplus \ldots \oplus \left(\alpha_k\beta_1 + \beta_k\right) \otimes \mathbf{v}_k\end{aligned}$$

Therefore any vector **u** in *V* is also a linear combination of $\mathbf{v}_2, \ldots, \mathbf{v}_k$, so $V = \operatorname{span}\{\mathbf{v}_2, \ldots, \mathbf{v}_k\}$.

**(b)** Suppose we are removing the vector $\mathbf{v}_r$ from *S* to obtain *T*. Since *S* is linearly independent, we cannot write $\mathbf{v}_r$ as a linear combination of the others.

Since *S* spans *V*, $\mathbf{v}_r \in V$. Since there is a vector in *V* that is not a linear combination of the vectors in *T*, *T* does not *V*.

Consider the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$. Suppose we want to check whether the set is linearly independent or not.

- Can we just check whether $\mathbf{v}_1$ is a linear combination of the other vectors, or must we check successively whether each of the vectors is a linear combination of the others?

**Theorem 6.2.1**

Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a set of vectors in a vector space. Then *S* is linearly independent if an only if the vector equation

$$k_1 \otimes \mathbf{v}_1 \oplus k_2 \otimes \mathbf{v}_2 \oplus \ldots \oplus k_r \otimes \mathbf{v}_r = \mathbf{0}$$

has only one solution, namely, the trivial solution $k_1 = k_2 = \ldots = k_r = 0$.

Equivalently, *S* is linearly dependent if and only if the vector equation has more than one solution, i.e. it has nontrivial solution where $k_1$, $k_2$, …, $k_r$ are not <u>all</u> zero.

*Proof:*

The equivalent statement is easier to prove.

Suppose $S$ is linearly dependent, we can find a vector in $S$, say $\mathbf{v}_1$, that can be written as a linear combination of the others in $S$, i.e.

$$\mathbf{v}_1 = \alpha_2 \otimes \mathbf{v}_2 \oplus \alpha_3 \otimes \mathbf{v}_3 \oplus ... \oplus \alpha_r \otimes \mathbf{v}_r \text{ for some scalars } \alpha_2, \alpha_3, ..., \alpha_r \in \mathbb{R}.$$

Then

$$(-1)\mathbf{v}_1 \oplus \alpha_2 \otimes \mathbf{v}_2 \oplus \alpha_3 \otimes \mathbf{v}_3 \oplus ... \oplus \alpha_r \otimes \mathbf{v}_r = \mathbf{0},$$

which shows the equation has solution $k_1 = -1, k_2 = \alpha_2, ..., k_r = \alpha_r$ are not all zero.

Conversely, suppose the equation

$$k_1 \otimes \mathbf{v}_1 \oplus k_2 \otimes \mathbf{v}_2 \oplus ... \oplus k_r \otimes \mathbf{v}_r = \mathbf{0}$$

has a solution where $k_1$ , $k_2$ , …, $k_r$ are not all zero, say $k_1 \neq 0$. Then

$$\mathbf{v}_1 = \left(-\frac{k_2}{k_1}\right) \otimes \mathbf{v}_2 \oplus ... \oplus \left(-\frac{k_r}{k_1}\right) \otimes \mathbf{v}_r .$$

Since one vector in $S$ is a linear combination of the other vectors in $S$, $S$ is linearly dependent.

**Example 6.2.2**

Determine whether the set $\{(1,0,2),(2,1,0),(-1,3,2)\}$ is linearly independent under the usual addition and scalar multiplication.

*Solution:*

Consider the vector equation

$$k_1(1,0,2) + k_2(2,1,0) + k_3(-1,3,2) = (0,0,0).$$

Comparing the components, we have

$$\begin{aligned} k_1 \quad +2k_2 \; -k_3 &= 0 \\ k_2 \quad +3k_3 &= 0 \\ 2k_1 \qquad\quad +2k_3 &= 0 \end{aligned}$$

Using Gaussian elimination, the above system reduces to the following equivalent system

$$\begin{aligned} k_1 \quad +2k_2 \; -k_3 &= 0 \\ k_2 \quad +3k_3 &= 0 \\ k_3 &= 0 \end{aligned}$$

It follows that $k_1 = k_2 = k_3 = 0$ is the only solution. Hence the set is linearly independent.

**Example 6.2.3**

Is the set $\left\{\begin{pmatrix}1 & 1\\0 & 1\end{pmatrix},\begin{pmatrix}-1 & 0\\1 & 1\end{pmatrix},\begin{pmatrix}2 & 1\\1 & -1\end{pmatrix},\begin{pmatrix}1 & 2\\3 & 2\end{pmatrix}\right\}$ of the vectors in $\mathbf{M}_{2,2}(\mathbb{R})$ linearly independent?

*Solution:*

Consider the vector equation

$$k_1\begin{pmatrix}1 & 1\\0 & 1\end{pmatrix}+k_2\begin{pmatrix}-1 & 0\\1 & 1\end{pmatrix}+k_3\begin{pmatrix}2 & 1\\1 & -1\end{pmatrix}+k_4\begin{pmatrix}1 & 2\\3 & 2\end{pmatrix}=\begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}$$

This leads to the following linear system

$$\begin{array}{lllll} k_1 & -k_2 & +2k_3 & +k_4 & =0 \\ k_1 & & +k_3 & +2k_4 & =0 \\ & k_2 & +k_3 & +3k_4 & =0 \\ k_1 & +k_2 & -k_3 & +2k_4 & =0 \end{array}$$

using Gaussian elimination, the above system reduces to the following equivalent system:

$$\begin{array}{lllll} k_1 & -k_2 & +2k_3 & +k_4 & =0 \\ & k_2 & -k_3 & +k_4 & =0 \\ & & k_3 & +k_4 & =0 \end{array}$$

This is a homogeneous linear system with more unknowns than equations, therefore it has a nontrivial solution. For example, $k_4=1$, $k_3=-1$, $k_2=-2$, $k_1=-1$. Consequently, the vector equation has a nontrivial solution. Hence the set is linearly dependent.

## 6.3 Basis

Having defined the concepts of span and linear independence, we now introduce a very important concept for vector space.

**Definition**

Let $V$ be a vector space and let $S=\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ be a set of vectors in $V$. Then $S$ is called a ***basis*** for $V$ if it satisfies the following two conditions:

- $S$ is linearly independent,
- $S$ spans $V$.

In other word, a basic $B$ of a vector space $V$ is a "minimal" spanning set for $V$, in the sense that if we remove any vector from $B$, the resulting set is no longer a spanning set for $V$. For example, we say the set $S=\{(1,0,0),(0,1,0),(0,0,1)\}$ is a basis of $\mathbb{R}^3$.

In H2 FM syllabus, we only consider vector spaces that have finite number of vectors in a basis.

**Example 6.3.1**

Show that the set $T = \{(1,0,1),(0,1,0),(0,1,1)\}$ is also a basis of $\mathbb{R}^3$.

*Solution:*

The vectors in *T* are in $\mathbb{R}^3$.

We first show that *T* is linearly independent. Consider the equation
$$\alpha(1,0,1)+\beta(0,1,0)+\gamma(0,1,1)=(0,0,0).$$

We have $\alpha = 0$, $\beta+\gamma=0$ and $\alpha+\gamma=0$. It is clear that $\alpha=\beta=\gamma=0$ is the only solution to the equation. This shows *T* is linearly independent.

Next we show that *T* spans $\mathbb{R}^3$. Take an arbitrary vector $(a,b,c)$ in $\mathbb{R}^3$. Now consider the equation
$$k(1,0,1)+l(0,1,0)+m(0,1,1)=(a,b,c).$$

We have $k=a$, $l+m=b$ and $k+m=c$. It is clear that $k=a$, $l=b-c+a$ and $m=c-a$. Thus,
$$(a,b,c)=a(1,0,1)+(b-c+a)(0,1,0)+(c-a)(0,1,1).$$

Therefore every vector in $\mathbb{R}^3$ is a linear combination of the vectors in *T*. Hence *T* spans $\mathbb{R}^3$.

Since *T* is linearly independent and it spans $\mathbb{R}^3$, we conclude that *T* is a basis for $\mathbb{R}^3$.

The above example show that $\mathbb{R}^3$ has another basis. In fact, $\mathbb{R}^3$ has many different bases. Among the bases of $\mathbb{R}^3$, the particular basis $\{(1,0,0),(0,1,0),(0,0,1)\}$ is called the ***standard basis*** of $\mathbb{R}^3$. The ***standard basis*** of $\mathbb{R}^n$ is defined in a similar way.

The ***standard basis*** of $\mathbf{P}_2$ is defined to be $\{1,x,x^2\}$. The ***standard basis*** of $\mathbf{M}_{2,2}(\mathbb{R})$ is defined to be $\left\{\begin{pmatrix}1&0\\0&0\end{pmatrix},\begin{pmatrix}0&1\\0&0\end{pmatrix},\begin{pmatrix}0&0\\1&0\end{pmatrix},\begin{pmatrix}0&0\\0&1\end{pmatrix}\right\}$. The standard bases of $\mathbf{P}_n$ and $\mathbf{M}_{x,y}(\mathbb{R})$ are defined in a similar way.

**Theorem 6.3.1**

Suppose $\{\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_n\}$ is a basis of a vector space *V*. Then every vector in *V* can be expressed uniquely as a linear combination of $\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_n$. Uniqueness here means that if $\mathbf{v}\in V$ and

$$\mathbf{v}=k_1\otimes\mathbf{v}_1\oplus k_2\otimes\mathbf{v}_2\oplus\dots\oplus k_n\otimes\mathbf{v}_n=c_1\otimes\mathbf{v}_1\oplus c_2\otimes\mathbf{v}_2\oplus\dots\oplus c_n\otimes\mathbf{v}_n,$$

then $k_1=c_1$, $k_2=c_2$, …, $k_n=c_n$.

*Proof:*

Let $k_1 \otimes \mathbf{v}_1 \oplus k_2 \otimes \mathbf{v}_2 \oplus \ldots \oplus k_n \otimes \mathbf{v}_n = c_1 \otimes \mathbf{v}_1 \oplus c_2 \otimes \mathbf{v}_2 \oplus \ldots \oplus c_n \otimes \mathbf{v}_n$.

By **Lemma 5.2.1**, we can 'subtract' $c_1 \otimes \mathbf{v}_1 \oplus c_2 \otimes \mathbf{v}_2 \oplus \ldots \oplus c_n \otimes \mathbf{v}_n$ from both sides (in fact, we are adding $(-c_1) \otimes \mathbf{v}_1 \oplus (-c_2) \otimes \mathbf{v}_2 \oplus \ldots \oplus (-c_n) \otimes \mathbf{v}_n$ to both sides) to obtain

$$(k_1 - c_1) \otimes \mathbf{v}_1 \oplus (k_2 - c_2) \otimes \mathbf{v}_2 \oplus \ldots \oplus (k_n - c_n) \otimes \mathbf{v}_n = \mathbf{0}.$$

As a basis of *V*, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is linearly independent. By **Theorem 6.2.1**, we have

$$k_1 - c_1 = k_2 - c_2 = \ldots = k_n - c_n,$$

i.e. $k_1 = c_1$, $k_2 = c_2$, …, $k_n = c_n$.

**Example 6.3.2**

Is $S = \{(1,0),(0,1),(1,-2)\}$ a basis for $\mathbb{R}^2$? Justify your answer.

*Solution:*

No. Since $(1,-2)$ is a linear combination of the other two vectors, i.e. $(1,-2) = 1(1,0) + (-2)(0,1)$, *S* is not linearly independent thus not a basis for $\mathbb{R}^2$.

**Example 6.3.3**

Is $S = \{(1,-1,0,0),(0,0,1,1)\}$ a basis for the vector space $V = \{(a,b,c,d) : a+b=0, c-d=0\}$? Justify your answer.

*Solution:*

Yes.

First note that $(1,-1,0,0)$ and $(0,0,1,1)$ are both in *V*.

It is clear that $(1,-1,0,0)$ is not a scalar multiple of $(0,0,1,1)$, i.e. $(1,-1,0,0) = k(0,0,1,1)$ has no solution, so *S* is linearly independent.

Now we need to show *S* spans *V*. Let $(a,b,c,d)$ be an arbitrary vector in *V*. Then $a+b=0$ and $c-d=0$. Consider the equation

$$(a,b,c,d) = k(1,-1,0,0) + l(0,0,1,1).$$

This equation gives $k = a = -b$ and $l = c = d$. Thus

$$(a,b,c,d) = a(1,-1,0,0) + c(0,0,1,1).$$

This shows that *S* spans *V*.
Since *S* is linearly independent and *S* spans *V*. *S* is a basis for *V*.

## 6.4 Dimension

**Theorem 6.4.1**

Let $V$ be a vector space and let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be a basis of $V$.

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a set of $k$ vectors <u>in</u> $V$.

**(a)** If $k > n$, then $S$ is linearly dependent.
**(b)** If $k < n$, then $S$ does not span $V$.

This theorem leads to the following:

**Theorem 6.4.2**

Suppose $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ and $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ are bases of a vector space $V$. Then $n = k$. In other words, all bases of a vector space have the <u>same number</u> of vectors.

Since all the bases of a vector space have the same number of vectors, we can make the following definition.

**Definition**

The ***dimension*** of a vector space $V$, denoted by $\dim(V)$, is defined to be the number of vectors in any basis of $V$. If the dimension of $V$ is finite, we say that $V$ is ***finite dimensional***.

We define the ***dimension*** of the zero vector space $\{\mathbf{0}\}$ as 0, with $\varnothing$ as its basis.

**Example 6.4.1**

What are the dimensions of $\mathbb{R}^n$, $\mathbf{P}_n$ and $\mathbf{M}_{m,n}(\mathbb{R})$?

*Solution:*

$\dim(\mathbb{R}^n) = n$ . $\dim(\mathbf{P}_n) = n+1$. $\dim(\mathbf{M}_{m,n}(\mathbb{R})) = mn$.

**Example 6.4.2**

Find the dimension of the subspace $\{(a,b,c,d) : a+b=0, c-d=0\}$ of $\mathbb{R}^4$.

*Solution:*

From **Example 6.3.3**, we know its dimension is 2.

Now we can rewrite **Theorem 6.4.1** using dimension,

**Theorem 6.4.3**

Let $V$ be a vector space with $\dim(V) = n > 0$

Let $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a set of $k$ vectors <u>in</u> $V$.

**(a)** If $k > n$, then $S$ is linearly dependent.
**(b)** If $k < n$, then $S$ does not span $V$.

- In **Theorem 6.4.3**, is it true that $S$ is linearly independent if $k \le n$?
- In **Theorem 6.4.3**, is it true that $S$ spans $V$ if $k \ge n$?
- In **Theorem 6.4.3**, is it true that $S$ is a basis for $V$ if $k = n$?

**Theorem 6.4.3** says that the <u>minimum</u> number of vectors needed to span $V$ is $\dim(V)$, and the <u>maximum</u> number of vectors in $V$ that are linearly independent is $\dim(V)$.

**Example 6.4.3**

Let $S = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ be a set of vectors in a vector space $V$. In each case below, can you say anything about $\dim(V)$?

**(a)** $S$ spans $V$.
**(b)** $S$ does not span $V$.
**(c)** $S$ is linearly independent.
**(d)** $S$ is linearly dependent.

*Solution:*

**(a)** $\dim(V) \le 3$.
**(b)** Nil.
**(c)** $\dim(V) \ge 3$.
**(d)** Nil.

**Theorem 6.4.4**

Let $V$ be a vector space with $\dim(V) = n > 0$. Let $S$ be a set of vectors <u>in</u> $V$ with <u>exactly</u> $n$ vectors. Then $S$ is a basis for $V$ if either $S$ spans $V$ or $S$ is linearly independent. That is

**(a)** If $S$ spans $V$, then $S$ is linearly independent.
**(b)** If $S$ is linearly independent, then $S$ spans $V$.

**Example 6.4.4**

Prove that the set $S = \{1+x+x^2, 1+2x+3x^2, x\}$ is a basis of $\mathbf{P}_2$.

*Proof:*

Since $\dim(\mathbf{P}_2) = 3$, we only need to verify one of the following:

**(a)** $S$ spans $\mathbf{P}_2$,

**(b)** $S$ is linearly independent.

We may choose to prove **(b)** in this proof.

Consider the vector equation:

$$\alpha(1+x+x^2) + \beta(1+2x+3x^2) + \gamma(x) = 0 + 0x + 0x^2 .$$

This leads to the homogenous linear system

$$\begin{aligned} \alpha \quad +\beta \quad +\gamma \quad &= 0 \\ \alpha \quad +2\beta \quad +3\gamma \quad &= 0 \\ \beta \quad &= 0 \end{aligned}$$

Solving the system by using its the augmented matrix, $\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$, we obtain

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 2 & 3 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Thus $\alpha = \beta = \gamma = 0$ is the only solution. Since $\dim(\mathbf{P}_2) = 3$ and $S$ is linearly independent, by **Theorem 6.4.4**, $S$ is a basis of $\mathbf{P}_2$.

- Can you use **Theorem 6.4.4** to construct another proof?

- Can you use the definition of *basis* to construct another proof?

**Theorem 6.4.5**

Let $V$ be a nonzero vector space.

**(a)** Every set of linearly independent vectors in $V$ can be enlarged to a basis of $V$, if necessary.
**(b)** Every spanning set of $V$ can be reduced to a basis of $V$, if necessary.

**Example 6.4.5**

Find a basis of $\mathbb{R}^3$ that contains the vector $(1,2,1)$.

*Solution:*

Since the set $\{(1,2,1)\}$ is linearly independent, it can be enlarged to a basis of $\mathbb{R}^3$ by **Theorem 6.4.5**. Since $\dim(\mathbb{R}^3)=3$, any basis of $\mathbb{R}^3$ has 3 vectors. We choose two standard basis vectors of $\mathbb{R}^3$, and consider the set $\{(1,2,1),(1,0,0),(0,1,0)\}$. The vector equation

$$k(1,2,1)+l(1,0,0)+m(0,1,0)=(0,0,0)$$

is equivalent to $k+l=0$, $2k+m=0$ and $k=0$, which obviously has only one solution $k=l=m=0$.

Thus $\{(1,2,1),(1,0,0),(0,1,0)\}$ is linearly independent and by **Theorem 6.4.4**, it is a basis for $\mathbb{R}^3$.

This example illustrates how we can possibly enlarge a set of linearly independent vectors in $V$ to a basis. We will discuss how to reduce a spanning set to a basis in **Section 7**.

The following theorem gives a relationship between the dimensions of a vector space and its subspaces.

**Theorem 6.4.6**

If $W$ is a *subspace* if a vector space $V$, then

$$\dim(W)\le\dim(V).$$

Furthermore, $\dim(W)=\dim(V)$ if and only if $W=V$.

**Example 6.4.6**

Let $W=\{a+bx+cx^2 : a-b+c=0\}$. Prove, without finding a basis for $W$, that $\dim(W)<3$.

*Proof:*

Since $W$ is a subspace of $\mathbf{P}_2$, $\dim(W)\le\dim(\mathbf{P}_2)=3$.
Observe that $1+x+x^2\in\mathbf{P}_2$ but $1+x+x^2\notin W$ as $1-1+1\ne 0$. $\dim(W)\ne\dim(\mathbf{P}_2)$.
Therefore $\dim(W)<3$.

## §7 Row Space, Column Space and Null Space

In this section, we define three vector spaces associated with a matrix. This will lead to the important concept of rank of a matrix, which has connection with the solution of a system of linear equations.

### 7.1 Row Space and Column Space

**Definition**

Let **A** be the $m \times n$ matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \ (1)$$

Let the rows of **A**, which are vectors in $\mathbb{R}^n$, be denoted by

$$\mathbf{r}_1 = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}$$
$$\mathbf{r}_1 = \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}$$
$$\vdots$$
$$\mathbf{r}_m = \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

and the columns of **A**, which are vectors in $\mathbb{R}^m$, be denoted by

$$\mathbf{c}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \mathbf{c}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \cdots, \mathbf{c}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

**(a)** The ***row space*** of **A** is defined to be the subspace of $\mathbb{R}^n$ spanned by the rows of **A**, i.e. row space of $\mathbf{A} = \text{span}\{\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_m\} = \{k_1\mathbf{r}_1 + k_2\mathbf{r}_2 + k_m\mathbf{r}_m : k_1, k_2, ..., k_m \in \mathbb{R}\}$.

**(b)** The ***column space*** of **A** is defined to be the subspace of $\mathbb{R}^m$ spanned by the columns of **A**, i.e. column space of $\mathbf{A} = \text{span}\{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n\} = \{k_1\mathbf{c}_1 + k_2\mathbf{c}_2 + k_n\mathbf{c}_n : k_1, k_2, ..., k_n \in \mathbb{R}\}$.

**Example 7.1.1**

Write down the row space and column space of $\mathbf{A} = \begin{pmatrix} 1 & 1 & -1 & 3 \\ -1 & 4 & 5 & -2 \\ 1 & 6 & 3 & 4 \end{pmatrix}$.

*Solution:*

$$\begin{aligned} \text{row space of } \mathbf{A} &= \text{span}\left\{\begin{pmatrix} 1 & 1 & -1 & 3 \end{pmatrix}, \begin{pmatrix} -1 & 4 & 5 & -2 \end{pmatrix}, \begin{pmatrix} 1 & 6 & 3 & 4 \end{pmatrix}\right\} \\ &= \left\{k_1\begin{pmatrix} 1 & 1 & -1 & 3 \end{pmatrix} + k_2\begin{pmatrix} -1 & 4 & 5 & -2 \end{pmatrix} + k_3\begin{pmatrix} 1 & 6 & 3 & 4 \end{pmatrix} : k_1, k_2, k_3 \in \mathbb{R}\right\} \end{aligned}$$

$$\text{column space of } \mathbf{A} = \text{span}\left\{ \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \\ 3 \end{pmatrix}, \begin{pmatrix} 3 \\ -2 \\ 4 \end{pmatrix} \right\}$$

$$= \left\{ k_1 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} + k_2 \begin{pmatrix} 1 \\ 4 \\ 1 \end{pmatrix} + k_3 \begin{pmatrix} -1 \\ 5 \\ 3 \end{pmatrix} + k_4 \begin{pmatrix} 3 \\ -2 \\ 4 \end{pmatrix} : k_1, k_2, k_3, k_4 \in \mathbb{R} \right\}$$

**Example 7.1.2**

Determine whether the vectors $\begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ in the column space of **A**, where $\mathbf{A} = \begin{pmatrix} 1 & -2 & -1 \\ 2 & -1 & 3 \\ 7 & -8 & 3 \end{pmatrix}$.

*Solution:*

Let $\begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix} = k_1 \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix} + k_2 \begin{pmatrix} -2 \\ -1 \\ -8 \end{pmatrix} + k_3 \begin{pmatrix} -1 \\ 3 \\ 3 \end{pmatrix}$.

Solving it, we have a solution $k_1 = 1$, $k_2 = 0$, $k_3 = -1$. Thus $\begin{pmatrix} 2 \\ -1 \\ 4 \end{pmatrix}$ is in the column space of **A**.

Let $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = l_1 \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix} + l_2 \begin{pmatrix} -2 \\ -1 \\ -8 \end{pmatrix} + l_3 \begin{pmatrix} -1 \\ 3 \\ 3 \end{pmatrix}$.

Solving it, we have no solution. Thus $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is not in the column space of **A**.

## 7.2 Null Space

**Definition**

Let **A** be the $m \times n$ matrix in (1). The set of all solutions of the homogeneous linear system $\mathbf{Ax} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$, called the ***null space*** of **A**, i.e.

$$\text{null space of } \mathbf{A} = \left\{ \mathbf{x} \in \mathbb{R}^n : \mathbf{Ax} = \mathbf{0} \right\} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} : \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right\}.$$

**Example 7.2.1**

Find the null space of **A**, where $\mathbf{A}=\begin{pmatrix}1 & -2 & -1\\ 2 & -1 & 3\\ 7 & -8 & 3\end{pmatrix}$. Hence, write down a basis for the null space.

*Solution:*

To find the null space of **A**, we solve the homogenous linear system $\mathbf{Ax}=\mathbf{0}$. The augmented matrix of the linear system is $\begin{pmatrix}1 & -2 & -1 & 0\\ 2 & -1 & 3 & 0\\ 7 & -8 & 3 & 0\end{pmatrix}$.

Reduce this to the reduced row-echelon form $\begin{pmatrix}1 & 0 & \frac{7}{3} & 0\\ 0 & 1 & \frac{5}{3} & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$, which corresponds to the linear system $x_1+\dfrac{7}{3}x_3=0$, $x_2+\dfrac{5}{3}x_3=0$.

Let $x_3=t$. Then the general solution to the linear system is $x_1=-\dfrac{7}{3}t$, $x_2=-\dfrac{5}{3}t$, $x_3=t$. Therefore the null space of **A** is

$$\left\{\begin{pmatrix}-\frac{7}{3}t\\ -\frac{5}{3}t\\ t\end{pmatrix}: t\in\mathbb{R}\right\}.$$

Note that every vector in the null space of **A** is of the form $t\begin{pmatrix}-\frac{7}{3}\\ -\frac{5}{3}\\ 1\end{pmatrix}$. Therefore the null space **A** is spanned by the set $S=\left\{\begin{pmatrix}-\frac{7}{3}\\ -\frac{5}{3}\\ 1\end{pmatrix}\right\}$. Since the vector in $S$ is nonzero, $S$ is linearly independent. Hence $S$ is a basis for the null space of **A**.

## 7.3 Finding Bases

The following theorems give a method for finding a basis for the row space of a matrix.

**Theorem 7.3.1**

Let **A** and **B** be matrices. If **B** can be obtained from **A** by performing a sequence of elementary row operations, then **A** and **B** have the same row space.

**Theorem 7.3.2**

If **R** is a matrix in row-echelon form, then the rows that containing the leading 1's form a basis for the row space of **R**.

**Example 7.3.1**

State a basis for the row space of $\mathbf{R}$, where $\mathbf{R}=\begin{pmatrix}1&2&1&0&3\\0&0&1&1&5\\0&0&0&0&1\\0&0&0&0&0\end{pmatrix}$.

*Solution:*

Since the matrix $\mathbf{R}$ is in row-echelon form, the set

$$\left\{\begin{pmatrix}1&2&1&0&3\end{pmatrix},\begin{pmatrix}0&0&1&1&5\end{pmatrix},\begin{pmatrix}0&0&0&0&1\end{pmatrix}\right\}$$

is a basis for the row space of $\mathbf{R}$.

**Example 7.3.2**

Find a basis for the row space of $\mathbf{B}$, where $\mathbf{B}=\begin{pmatrix}1&-2&0&0&3\\2&-5&-3&-2&6\\0&5&15&10&0\\2&6&18&8&6\end{pmatrix}$.

*Solution:*

We perform elementary row operations to reduce $\mathbf{B}$ to row-echelon form: $\begin{pmatrix}1&-2&0&0&3\\0&1&3&2&0\\0&0&1&1&0\\0&0&0&0&0\end{pmatrix}$.

Therefore, the set $\left\{\begin{pmatrix}1&-2&0&0&3\end{pmatrix},\begin{pmatrix}0&1&3&2&0\end{pmatrix},\begin{pmatrix}0&0&1&1&0\end{pmatrix}\right\}$ is a basis of $\mathbf{B}$.

The following theorems give a method to find a basis for the column space of a matrix.

**Theorem 7.3.3**

If $\mathbf{R}$ is a matrix in row-echelon form, then the columns that containing the leading 1's form a basis for the column space of $\mathbf{R}$.

**Theorem 7.3.4**

Let $\mathbf{A}$ and $\mathbf{B}$ be matrices. Suppose $\mathbf{B}$ can be obtained be obtained from $\mathbf{A}$ by performing a sequence of elementary row operations. Then a given set of columns of $\mathbf{A}$ form a basis for the column space of $\mathbf{A}$ if and only if the *corresponding columns* of $\mathbf{B}$ form a basis for the column space of $\mathbf{B}$.

**Example 7.3.3**

State a basis for the column space of $\mathbf{R}$, where $\mathbf{R}=\begin{pmatrix}1&2&1&0&3\\0&0&1&1&5\\0&0&0&0&1\\0&0&0&0&0\end{pmatrix}$.

*Solution:*

Since the matrix **R** is in row-echelon form, $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a basis for the column space of **R**.

**Example 7.3.4**

It is given that $\mathbf{B} = \begin{pmatrix} 1 & -2 & 0 & 0 & 3 \\ 2 & -5 & -3 & -2 & 6 \\ 0 & 5 & 15 & 10 & 0 \\ 2 & 6 & 18 & 8 & 6 \end{pmatrix}$ can be reduced to row-echelon form $\begin{pmatrix} 1 & -2 & 0 & 0 & 3 \\ 0 & 1 & 3 & 2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$.
State a basis for the column space of **B**.

*Solution:*

Since the leading 1's in **B**'s row-echelon form are in the 1st, 2nd and 3rd columns, a basis for the column space of **B** is $\left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ -5 \\ 5 \\ 6 \end{pmatrix}, \begin{pmatrix} 0 \\ -3 \\ 15 \\ 18 \end{pmatrix} \right\}$.

- Is $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 1 \\ 0 \end{pmatrix} \right\}$ also a basis for the column space of **B**?

- Can you suggest another possible method to find a basis for the column space of a given matrix?

Now we can apply the method for finding a basis for the column space of a matrix to reduce a spanning set for a subspace of $\mathbb{R}^n$ to a basis of that subspace.

**Example 7.3.5**

Let $W$ be the subspace of $\mathbb{R}^n$ spanned by the set

$$S = \left\{ \begin{pmatrix} 1 \\ -2 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ -5 \\ -3 \\ 6 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 4 \\ 7 \end{pmatrix}, \begin{pmatrix} 5 \\ -8 \\ 1 \\ 16 \end{pmatrix} \right\}.$$

Reduce $S$ to a basis of $W$.

*Solution:*

Construct a matrix **A** whose columns are the vectors in *S*: $\mathbf{A}=\begin{pmatrix}1&2&0&2&5\\-2&-5&1&-1&-8\\0&-3&3&4&1\\3&6&0&7&16\end{pmatrix}$.

Note that the column space of **A** is *W*. Therefore a basis for the column space of **A** is a basis for *W*.

Now reduce **A** to row-echelon form: $\begin{pmatrix}1&2&0&2&5\\0&1&-1&-3&-2\\0&0&0&1&1\\0&0&0&0&0\end{pmatrix}$.

Since the 1st, 2nd and 4th columns contain the leading 1's, the 1st, 2nd and 4th columns of **A** form a bsis for the column space of **A**.

Thus, $\left\{\begin{pmatrix}1\\-2\\0\\3\end{pmatrix},\begin{pmatrix}2\\-5\\-3\\6\end{pmatrix},\begin{pmatrix}2\\-1\\4\\7\end{pmatrix}\right\}$ is a basis for *W*.

## 7.4 Rank and Nullity

**Theorem 7.4.1**

For any matrix **A**, the <u>dimension</u> of the row space of **A** is equal to the <u>dimension</u> of the column space of **A**.

- How can we justify this theorem?

**Definition**

The common dimension of the row space and column space of a matrix **A** is called the ***rank*** of **A**, and is denoted by $\text{rank}(\mathbf{A})$.

The dimension of the null space of **A** is called the ***nullity*** of **A**, and is denoted by $\text{nullity}(\mathbf{A})$.

**Example 7.4.1**

It is given that $\mathbf{A}=\begin{pmatrix}1&-2&-1\\2&-1&3\\7&-8&3\end{pmatrix}$.

**(a)** Deduce the nullity of **A** from **Example 7.2.1**.

**(b)** Find the rank of **A**.

*Solution:*

**(a)** nullity$(\mathbf{A})=1$ as there is only one vector in the basis of its null space.

**(b)** Reduce $\mathbf{A}=\begin{pmatrix}1&-2&-1\\2&-1&3\\7&-8&3\end{pmatrix}$ to row-echelon form: $\mathbf{A}=\begin{pmatrix}1&-\frac{8}{7}&\frac{3}{7}\\0&1&\frac{5}{3}\\0&0&0\end{pmatrix}$. Since there are 2 leading 1's, rank$(\mathbf{A})=2$.

**Example 7.4.2**

It is given that $\mathbf{B}=\begin{pmatrix}1&-2&0&0&3\\2&-5&-3&-2&6\\0&5&15&10&0\\2&6&18&8&6\end{pmatrix}$.

**(a)** Deduce the rank of **B** from the result of **Example 7.3.4**.
**(b)** Find the nullity of **B**.

*Solution:*

**(a)** rank$(\mathbf{B})=3$ as its row-echelon form, $\begin{pmatrix}1&-2&0&0&3\\0&1&3&2&0\\0&0&1&1&0\\0&0&0&0&0\end{pmatrix}$, has 3 leading 1's.

**(b)** To find the null space of **B**, we solve the homogenous linear system $\mathbf{Bx}=\mathbf{0}$. The augmented matrix has row-echelon form $\begin{pmatrix}1&-2&0&0&3&0\\0&1&3&2&0&0\\0&0&1&1&0&0\\0&0&0&0&0&0\end{pmatrix}$, which corresponds to a linear system:

$$x_1-2x_2+3x_5=0,\ x_2+3x_3+2x_4=0,\ x_3+x_4=0.$$

Let $x_4=s$ and $x_5=t$.
We have $x_3=-x_4=-s$, $x_2=-3x_3-2x_4=3s-2s=s$, $x_1=2x_2-3x_5=2s-3t$.

Therefore, the null space of **B** is $\left\{\begin{pmatrix}2s-3t\\s\\-s\\s\\t\end{pmatrix}:s,t\in\mathbb{R}\right\}=\left\{s\begin{pmatrix}2\\1\\-1\\1\\0\end{pmatrix}+t\begin{pmatrix}-3\\0\\0\\0\\1\end{pmatrix}:s,t\in\mathbb{R}\right\}$. Since there are two vectors in its basis, nullity$(\mathbf{B})=3$.

If **A** is a matrix, then

rank$(\mathbf{A})$ = the number of leading 1's in the row-echelon form of **A**;

nullity$(\mathbf{A})$ = the number of parameters in the general solution of $\mathbf{Ax}=\mathbf{0}$.

- Can you make a conjecture for the relationship between the rank and nullity of a matrix

**Theorem 7.4.2**

If **A** is an $m \times n$ matrix, then

$$\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n.$$

The following theorem links the rank of **A** to the solutions of the linear system $\mathbf{Ax} = \mathbf{b}$.

**Theorem 7.4.3**

Let **A** be an $m \times n$ matrix. Then the following statements are equivalent.

**(a)** The linear system $\mathbf{Ax} = \mathbf{b}$ is consistent.
**(b)** The rank of **A** is equal to the rank of the augmented matrix $(\mathbf{A} \mid \mathbf{b})$.
**(c)** **b** is in the column space of **A**.

**Example 7.4.2**

Consider the linear system $\mathbf{Ax} = \mathbf{b}$. What can you say about the relationship between the rank of the coefficient matrix and the rank of the augmented matrix, when the system is *inconsistent*?

*Solution:*

Let $\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n$ denote the columns of **A**, and let $B = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k\}$ be a basis of **A**. Then we know

$$\text{column space of } \mathbf{A} = \text{span}(B),$$

and $B$ is linearly independent, and $\text{rank}(\mathbf{A}) = \dim(\text{the column space of } \mathbf{A}) = k$.

Now consider the set of vectors, $B' = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k, \mathbf{b}\}$. All the vectors in $B'$ are in the column space of $(\mathbf{A} \mid \mathbf{b})$.

Since the system $\mathbf{Ax} = \mathbf{b}$ is given to be inconsistent, **b** is not in the column space of **A**. Thus **b** is not a linear combination of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$. This means the vector equation

$$l_1\mathbf{v}_1 + l_2\mathbf{v}_2 + l_k\mathbf{v}_k + l\,\mathbf{b} = \mathbf{0}$$

has no solution when $l \neq 0$. When $l = 0$, the only solution $l_1 = l_2 = ... = l_k = 0$ as $B$ is linearly independent. Thus $B'$ is linearly independent.

By definition, the column space of the augmented matrix $(\mathbf{A} \mid \mathbf{b})$ is

$$\text{span}\{\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n, \mathbf{b}\} = \{m_1\mathbf{c}_1 + m_2\mathbf{c}_2 + ... + m_n\mathbf{c}_n + m\mathbf{b} : m_1, m_2, ..., m_n, m \in \mathbb{R}\}.$$

Since each of $\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_k$, we can always rewrite $m_1\mathbf{c}_1 + m_2\mathbf{c}_2 + ... + m_n\mathbf{c}_n + m\mathbf{b}$ in the form $l_1\mathbf{v}_1 + l_2\mathbf{v}_2 + ... + l_k\mathbf{v}_k + m\mathbf{b}$. Thus, $B'$ also spans $(\mathbf{A} \mid \mathbf{b})$.

Since the vectors in $B'$ are in the column space of $(\mathbf{A} \mid \mathbf{b})$, $B'$ is linearly independent and $B'$ spans $(\mathbf{A} \mid \mathbf{b})$, $B'$ is a basis for the column space of $(\mathbf{A} \mid \mathbf{b})$. Thus $\text{rank}(\mathbf{A} \mid \mathbf{b}) = k + 1 = \text{rank}(\mathbf{A}) + 1$.

**Example 7.4.3**

Consider the matrix equation $\mathbf{Au} = \mathbf{b}$ that corresponds to the following system of two linear equations and two unknowns:

$$\begin{aligned} a_{11}x + a_{12}y &= b_1 \\ a_{21}x + a_{22}y &= b_2 \end{aligned} \quad (*)$$

Let the rank of its coefficient matrix be $r$ and the rank of its augmented matrix be $q$. It is assumed that neither row of the coefficient matrix contains only 0.

**(i)** Find all the possible values for the ordered pair $(r, q)$.

Geometrically, each equation in (*) represents a line on 2-dimensional plane.

**(ii)** What can you say about the relationship between the values of $(r, q)$ and the intersection of the two lines?

*Solution:*

**(i)** Since there are only 2 rows in both the coefficient matrix and the augmented matrix, and no row contains only 0's, we can say $1 \le r, q \le 2$.

We also know that $r = q$ when the system is consistent and $r + 1 = q$ when the system is not.

Thus $(r, q) = (1, 1)$ or $(r, q) = (1, 2)$ or $(r, q) = (2, 2)$.

**(ii)** Case 1: $(r, q) = (2, 2)$.
This implies the system is consistent. The leading 1's occur in the first and the second columns of the row-echelon form of the augmented matrix $(\mathbf{A} \mid \mathbf{b})$, i.e.

$$\left(\begin{array}{cc|c} 1 & a_{21}' & b_1' \\ 0 & 1 & b_2' \end{array}\right).$$

It has a unique solution, $y = b_2'$ and $x = b_1' - a_{21}'b_2'$.
In this case, the lines intersect at a point.

Case 2: $(r, q) = (1, 2)$.
This implies the system is inconsistent, i.e. it has no solution.
In this case, the lines are parallel.

Case 3: $(r, q) = (1, 1)$.
This implies the system is consistent. There is only one leading 1 in the row-echelon form of the augmented matrix, so the second row contains only 0's. The first row of the row-echelon form represents the equation of a line.
In this case, the lines coincides (or overlaps).

## §8 Linear Transformations

In H2 Mathematics, we have learnt how to write descriptions for certain transformations of graphs, but these graph transformations can be quantified! With linear transformations, we can quantify many graph transformations such as reflections, scaling, shears and rotations on 2-D plane or even in 3-D space. You may refer to **Appendix III** for more details.

### 8.1 Linear Transformations in General

**Definition**

If $V$ and $W$ are vector spaces, then a ***linear transformation*** (also called ***linear map*** *or* ***linear mapping***) is a function $\mathrm{T}: V \to W$ that preserves the operations of addition and scalar multiplications, i.e. for all vectors **u** and **v** in $V$ and all scalars $k$:

$$\mathrm{T}(\mathbf{u} \oplus \mathbf{v}) = \mathrm{T}(\mathbf{u}) \oplus \mathrm{T}(\mathbf{v}) \text{ and } \mathrm{T}(k \otimes \mathbf{u}) = k \otimes \mathrm{T}(\mathbf{u}).$$

Note that the addition and scalar multiplication on the left-hand side are defined for the vector space $V$, and those on the right-hand side are defined for the vector space $W$. They need not be the same in general.

**Example 8.1.1**

Prove that $\mathrm{L}: \mathbb{R} \to \mathbb{R}$ is a linear transformation if $\mathrm{L}(x) = 2x$.

*Proof:*

Consider $x_1, x_2, k \in \mathbb{R}$.
$\mathrm{L}(x_1) = 2x_1$, $\mathrm{L}(x_2) = 2x_2$, $\mathrm{L}(x_1 + x_2) = 2(x_1 + x_2) = 2x_1 + 2x_2 = \mathrm{L}(x_1) + \mathrm{L}(x_2)$.
$\mathrm{L}(kx_1) = 2(kx_1) = k(2x_1) = k\mathrm{L}(x_1)$.
Thus, L is a linear transformation.

**Example 8.1.2**

Determine whether each of the following function is a linear transformation. Justify your answers.

**(a)** $\mathrm{T}_1: \mathbb{R} \to \mathbb{R}$, $\mathrm{T}_1(x) = 2x + 1$.

**(b)** $\mathrm{T}_2: \mathbb{R}^2 \to \mathbb{R}^2$, $\mathrm{T}_2(\mathbf{u}) = \mathbf{Au}$ where $\mathbf{A}$ is a fixed $2 \times 2$ matrix.

**(c)** $\mathrm{T}_3: \mathbf{P}_n \to \mathbf{P}_{n-1}$ ($n \geq 1$), $\mathrm{T}_3(p(x)) = p'(x)$.

**(d)** $\mathrm{T}_4: \mathbb{R} \to \mathbb{R}$, $\mathrm{T}_4(\theta) = \sin\theta$.

**(e)** $\mathrm{T}_5: \mathbb{R}^2 \to \mathbb{R}^3$, $\mathrm{T}_5\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x + y \\ x - y + 1 \\ y \end{pmatrix}$.

**(f)** $\mathrm{T}_6: \mathbb{R}^3 \to \mathbf{P}_2$, $\mathrm{T}_6(a, b, c) = ax^2 + bx + c$.

*Solution:*

**(a)** No. $\mathrm{T}_1(1)=3$, $\mathrm{T}_1(2)=5$ but $\mathrm{T}_1(1+2)=\mathrm{T}_1(3)=7\neq 8=\mathrm{T}_1(1)+\mathrm{T}_1(2)$.

**(b)** Yes. For all vectors **u** and **v** in $\mathbb{R}^2$ and all scalars *k*,
$\mathrm{T}_2(\mathbf{u}+\mathbf{v})=\mathbf{A}(\mathbf{u}+\mathbf{v})=\mathbf{Au}+\mathbf{Av}=\mathrm{T}_2(\mathbf{u})+\mathrm{T}_2(\mathbf{v})$, $\mathrm{T}(k\mathbf{u})=\mathbf{A}(k\mathbf{u})=k(\mathbf{Au})=k\mathrm{T}(\mathbf{u})$.

**(c)** Yes. For all vectors $p(x)$ and $q(x)$ in $\mathbf{P}_n$ and all scalars *k*,

$$\mathrm{T}_3\left(p(x)+q(x)\right)=\frac{\mathrm{d}}{\mathrm{d}x}\left(p(x)+q(x)\right)=p'(x)+q'(x)=\mathrm{T}_3\left(p(x)\right)+\mathrm{T}_3\left(q(x)\right),$$

$$\mathrm{T}_3\left(kp(x)\right)=\frac{\mathrm{d}}{\mathrm{d}x}\left(kp(x)\right)=kp'(x)=k\mathrm{T}_3\left(p(x)\right).$$

**(d)** No. $\mathrm{T}_4\left(\frac{\pi}{2}\right)=\sin\frac{\pi}{2}=1$ but $\mathrm{T}_4\left(2\times\frac{\pi}{2}\right)=\mathrm{T}_4(\pi)=\sin\pi=0\neq 2=2\mathrm{T}_4\left(\frac{\pi}{2}\right)$.

**(e)** No. $\mathrm{T}_5\begin{pmatrix}0\\0\end{pmatrix}=\begin{pmatrix}0\\1\\0\end{pmatrix}$, $\mathrm{T}_5\begin{pmatrix}0\\1\end{pmatrix}=\begin{pmatrix}1\\0\\1\end{pmatrix}$, $\mathrm{T}_5\left[\begin{pmatrix}0\\1\end{pmatrix}+\begin{pmatrix}0\\0\end{pmatrix}\right]=\mathrm{T}_5\left[\begin{pmatrix}0\\1\end{pmatrix}\right]=\begin{pmatrix}1\\0\\1\end{pmatrix}\neq\begin{pmatrix}1\\1\\1\end{pmatrix}=\mathrm{T}_5\left[\begin{pmatrix}0\\0\end{pmatrix}\right]+\mathrm{T}_5\left[\begin{pmatrix}0\\1\end{pmatrix}\right]$.

**(f)** Yes. For all vectors $(a_1,b_1,c_1)$ and $(a_2,b_2,c_2)$ in $\mathbb{R}^3$ and all scalars *k*,

$$\mathrm{T}_6\left((a_1,b_1,c_1)+(a_2,b_2,c_2)\right)=\mathrm{T}_6\left(a_1+a_2,b_1+b_2,c_1+c_2\right)=(a_1+a_2)x^2+(b_1+b_2)x+(c_1+c_2),$$

$$\mathrm{T}_6(a_1,b_1,c_1)+\mathrm{T}_6(a_2,b_2,c_2)=\left(a_1x^2+b_1x+c_1\right)+\left(a_2x^2+b_2x+c_2\right)=\mathrm{T}_6(a_1,b_1,c_1)+\mathrm{T}_6(a_2,b_2,c_2).$$

$$\mathrm{T}_6\left(k(a_1,b_1,c_1)\right)=\mathrm{T}_6(ka_1,kb_1,kc_1)=(ka_1)x^2+(kb_1)x+(kc_1)=k\left(a_1x^2+b_1x+c_1\right)=k\mathrm{T}_6(a_1,b_1,c_1)$$

**Theorem 8.1.1**

If $\mathrm{T}:V\to W$ is a linear transformation, then $\mathrm{T}(\mathbf{0})=\mathbf{0}$.

- Are the **0** inside the brackets the same as the **0** on the right-hand side?

*Proof:*

$\mathrm{T}(\mathbf{0})\oplus\mathbf{0}=\mathrm{T}(\mathbf{0})=\mathrm{T}(\mathbf{0}\oplus\mathbf{0})=\mathrm{T}(\mathbf{0})\oplus\mathrm{T}(\mathbf{0})$. Adding $-\mathrm{T}(\mathbf{0})$ to both sides (or by **Lemma 5.2.1**), we have $\mathrm{T}(\mathbf{0})=\mathbf{0}$.

**Theorem 8.1.2**

If $\mathrm{T}: V \rightarrow W$ is a linear transformation, then

$$\mathrm{T}(a \otimes \mathbf{u} \oplus b \otimes \mathbf{v}) = a \otimes \mathrm{T}(\mathbf{u}) \oplus b \otimes \mathrm{T}(\mathbf{v}) \text{ for all } \mathbf{u}, \mathbf{v} \in V \text{ and } a, b \in \mathbb{R},$$

or more generally,

$$\mathrm{T}(k_1 \otimes \mathbf{v}_1 \oplus k_2 \otimes \mathbf{v}_2 \oplus \ldots \oplus k_n \otimes \mathbf{v}_n) = k_1 \otimes \mathrm{T}(\mathbf{v}_1) \oplus k_2 \otimes \mathrm{T}(\mathbf{v}_2) \oplus \ldots \oplus k_n \otimes \mathrm{T}(\mathbf{v}_n)$$

for all $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ and $k_1, k_2, \ldots, k_n \in \mathbb{R}$.

*Proof:*

$\mathrm{T}(a \otimes \mathbf{u} \oplus b \otimes \mathbf{v}) = \mathrm{T}(a \otimes \mathbf{u}) \oplus \mathrm{T}(b \otimes \mathbf{v}) = a \otimes \mathrm{T}(\mathbf{u}) \oplus b \otimes \mathrm{T}(\mathbf{v})$.
To prove the more general result, you may use mathematical induction.

Note if $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for *V*, then this linear transformation $\mathrm{T}: V \rightarrow W$ is uniquely determined by $\mathrm{T}(\mathbf{v}_1), \mathrm{T}(\mathbf{v}_2), \ldots, \mathrm{T}(\mathbf{v}_n)$.

Use Example **8.1.2(c)** to illustrate this point: since $\{1, x, \ldots, x^n\}$ is a basis for $\mathbf{P}_n$, as long as we know how to differentiate (or transform) these vectors, we know how to differentiate all the other vectors in $\mathbf{P}_n$.

## 8.2 Null Space and Range Space of Linear Transformation

**Definition**

Let $\mathrm{T}: V \rightarrow W$ be a linear transformation. Then

***null space*** of $\mathrm{T} = \{\mathbf{x} \in V : \mathrm{T}(\mathbf{x}) = \mathbf{0}\}$,
***range space*** of $\mathrm{T} = \{\mathrm{T}(\mathbf{x}) : \mathbf{x} \in V\}$.

- The null space of T is a subset of ___ and the range space of T is a subset of ___.

### Example 8.2.1

Let $\mathrm{T}: V \rightarrow W$ be a linear transformation, prove that

**(a)** the null space of T is vector space,

**(b)** the range space of T is also a vector space.

*Proof:*

**(a)** Since the null space is a subset of $V$, by **Theorem 5.3.1** we just need to show that it is nonempty and closed under addition and scalar multiplication.

The null space is nonempty as it contains **0**, since $\mathrm{T}(\mathbf{0})=\mathbf{0}$.

Let $\mathbf{u},\mathbf{v}\in$ the null space $\subseteq V$ and $k\in\mathbb{R}$. Then $\mathrm{T}(\mathbf{u})=\mathbf{0}$ and $\mathrm{T}(\mathbf{v})=\mathbf{0}$.
Since $\mathrm{T}(\mathbf{u}\oplus\mathbf{v})=\mathrm{T}(\mathbf{u})\oplus\mathrm{T}(\mathbf{v})=\mathbf{0}\oplus\mathbf{0}=\mathbf{0}$, $\mathbf{u}\oplus\mathbf{v}\in$ the null space.
Since $\mathrm{T}(k\otimes\mathbf{u})=k\otimes\mathrm{T}(\mathbf{u})=k\otimes\mathbf{0}=\mathbf{0}$, $k\otimes\mathbf{u}\in$ the null space.

Therefore, the null space is a vector space (a subspace of $V$).

**(b)** Since the range space is a subset of $W$, by **Theorem 5.3.1** we just need to show that it is nonempty and closed under addition and scalar multiplication.

The range space is non empty as it contains **0**, since $\mathrm{T}(\mathbf{0})=\mathbf{0}$.

Let $\mathbf{w},\mathbf{x}\in$ the range space $\subseteq W$ and $l\in\mathbb{R}$. Then $\mathrm{T}(\mathbf{p})=\mathbf{w}$ and $\mathrm{T}(\mathbf{q})=\mathbf{x}$ for some $\mathbf{p},\mathbf{q}\in V$.
Since $\mathbf{p}\oplus\mathbf{q}\in V$ and $\mathrm{T}(\mathbf{p}\oplus\mathbf{q})=\mathrm{T}(\mathbf{p})\oplus\mathrm{T}(\mathbf{q})=\mathbf{w}\oplus\mathbf{x}$, $\mathbf{w}\oplus\mathbf{x}\in$ the range space.
Since $l\otimes\mathbf{p}\in V$ and $\mathrm{T}(l\otimes\mathbf{p})=l\otimes\mathrm{T}(\mathbf{p})=l\otimes\mathbf{w}$, $l\otimes\mathbf{w}\in$ the range space.

Therefore, the range space is a vector space (a subset of $W$).

**Definition**

Let $\mathrm{T}:V\to W$ be a linear transformation. Then the ***rank*** of T is the dimension of the range space of T, and the ***nullity*** of T is the dimension of the null space of T.

**Example 8.2.2**

Find the rank and nullity of each of the following linear transformation:

**(a)** $\mathrm{L}:\mathbb{R}\to\mathbb{R}$, $\mathrm{L}(x)=2x$.
**(b)** $\mathrm{T}_6:\mathbb{R}^3\to\mathbf{P}_2$, $\mathrm{T}_6(a,b,c)=ax^2+bx+c$.
**(c)** $\mathrm{T}_3:\mathbf{P}_n\to\mathbf{P}_{n-1}$ ($n\geq1$), $\mathrm{T}_3(p(x))=p'(x)$.

*Solution:*

**(a)** Since the range space $\{2x:x\in\mathbb{R}\}=\mathbb{R}$, the range space has a basis $\{1\}$, so its dimension is 1. Thus, $\mathrm{rank}(\mathrm{L})=1$.

$\mathrm{L}(x)=0$ implies $x=0$, the null space has only a zero vector in it, so its dimension is 0. Thus, $\mathrm{nullity}(\mathrm{L})=1$.

**(b)** Since the range space $\{ax^2+bx+c:(a,b,c)\in\mathbb{R}^3\}=\mathbf{P}_2$, its dimension is 3. Thus $\mathrm{rank}(\mathrm{T}_6)=3$.

$T_6(a,b,c) = 0x^2 + 0x + 0$ implies $(a,b,c) = (0,0,0)$, the null space has only a zero vector in it, so its dimension is 0. Thus $\text{nullity}(T_6) = 3$.

**(c)** Since the range space $\{p'(x) : p(x) \in \mathbf{P}_n\} = \mathbf{P}_{n-1}$, its dimension is $n$. Thus $\text{rank}(T_3) = 3$.

$T_3(p(x)) = 0$ implies $p(x)$ can be any real number, so the null space is $\mathbf{P}_0$ with dimension 1. Thus $\text{nullity}(T_3) = 1$.

- What conjecture can you form about the rank and nullity of a linear transformation?
- What can you say about the rank and nullity of the linear transformation $T_2 : \mathbb{R}^2 \to \mathbb{R}^2$, $T_2(\mathbf{u}) = \mathbf{Au}$ where $\mathbf{A}$ is a fixed $2\times 2$ matrix.

## 8.3 Linear Transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$

We have seen some similarities between a linear transformation and a matrix. In this session, we shall discuss the similarities in details.

> **Theorem 8.3.1**
>
> Any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ can be represented by a $m \times n$ matrix **A**, such that
> $$T(\mathbf{x}) = \mathbf{Ax} \text{ for all } \mathbf{x} \in \mathbb{R}^n$$

Before proving this theorem, let us look at a few examples:

**Example 8.3.1**

Identify the matrices that represent the following linear transformations:

**(a)** $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+2y \\ x-y \end{pmatrix}$.

**(b)** $L : \mathbb{R}^3 \to \mathbb{R}^2$, $L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+\sqrt{2}y \\ 3x-0.5z \end{pmatrix}$.

*Solution:*

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+2y \\ x-y \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \quad L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+\sqrt{2}y \\ 3x-0.5z \end{pmatrix} = \begin{pmatrix} 1 & \sqrt{2} & 0 \\ 3 & 0 & -0.5 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

*Proof for* ***Theorem 8.3.1***:

Let the standard basis of $\mathbb{R}^n$ be $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n\}$, so any vector in $\mathbb{R}^n$ is a linear combination of these vectors, i.e.

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_n\mathbf{e}_n.$$

We construct a matrix $\mathbf{A}$, where $\mathbf{A} = \begin{pmatrix}\mathrm{T}(\mathbf{e}_1) & \mathrm{T}(\mathbf{e}_2) & ... & \mathrm{T}(\mathbf{e}_n)\end{pmatrix}$, each of the columns, $\mathrm{T}(\mathbf{e}_i)$ is a vector in $\mathbb{R}^m$, thus $\mathbf{A}$ is a $m \times n$ matrix, i.e.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Now,

$$\begin{aligned} \mathrm{T}(\mathbf{x}) &= \mathrm{T}(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_n\mathbf{e}_n) \\ &= x_1\mathrm{T}(\mathbf{e}_1) + x_2\mathrm{T}(\mathbf{e}_2) + ... + x_n\mathrm{T}(\mathbf{e}_n) \\ &= x_1\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2\begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + ... + x_n\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ a_{21}x_1 + a_{122}x_2 + ... + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \mathbf{Ax}. \end{aligned}$$

**Example 8.3.2**

Let $\mathrm{T}: \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with $\mathrm{T}\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1\\-1\end{pmatrix}$ and $\mathrm{T}\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}2\\3\end{pmatrix}$.

**(i)** Find $\mathrm{T}\begin{pmatrix}2\\1\end{pmatrix}$.

**(ii)** State the matrix $\mathbf{A}$ such that $\mathrm{T}(\mathbf{u}) = \mathbf{Au}$ for all $\mathbf{u} \in \mathbb{R}^2$.

*Solution:*

**(i)** $\mathrm{T}\begin{pmatrix}2\\1\end{pmatrix} = \mathrm{T}\left[2\begin{pmatrix}1\\0\end{pmatrix} + \begin{pmatrix}0\\1\end{pmatrix}\right] = 2\mathrm{T}\begin{pmatrix}1\\0\end{pmatrix} + \mathrm{T}\begin{pmatrix}0\\1\end{pmatrix} = 2\begin{pmatrix}1\\-1\end{pmatrix} + \begin{pmatrix}2\\3\end{pmatrix} = \begin{pmatrix}4\\1\end{pmatrix}$.

**(ii)** $\mathbf{A} = \begin{pmatrix}1 & 2\\-1 & 3\end{pmatrix}$.

**Theorem 8.3.2**

Let $\mathrm{T}:\mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and let $\mathbf{A}$ be the matrix representing T, i.e.

$$\mathrm{T}(\mathbf{x}) = \mathbf{Ax} \text{ for all } \mathbf{x} \in \mathbb{R}^n.$$

Then

$$\text{null space of T} = \text{null space of } \mathbf{A},$$
$$\text{range space of T} = \text{column space of } \mathbf{A},$$

and consequently,

$$\text{nullity}(\mathrm{T}) + \text{rank}(\mathrm{T}) = \dim(\mathbb{R}^n) = n.$$

**Example 8.3.3**

The linear transformation $\sigma:\mathbb{R}^3 \to \mathbb{R}^3$ is represented by the matrix

$$\begin{pmatrix} 2 & 1 & 4 \\ -1 & 3 & -9 \\ 3 & 1 & 7 \end{pmatrix}$$

with respect to the standard basis of $\mathbb{R}^3$.

**(i)** Show that the range space of $\sigma$ has dimension 2, and state the nullity of $\sigma$.

**(ii)** Given that $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is in the range space of $\sigma$, obtain integers *a*, *b*, *c*, not all zero, such that

$$ax + by + cz = 0.$$

**(iii)** Find the subset *P* of $\mathbb{R}^3$ whose image under $\sigma$ is the line

$$\mathbf{r} = \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix} + \lambda \begin{pmatrix} 4 \\ 5 \\ 5 \end{pmatrix}.$$

Show that *P* is a plane, and give its equation in the form $kx + ly + mz = n$, where *k*, *l*, *m*, *n* are integers.

*Solution:*

**(i)** Let $\mathbf{A} = \begin{pmatrix} 2 & 1 & 4 \\ -1 & 3 & -9 \\ 3 & 1 & 7 \end{pmatrix}$. Performing elementary row operations on **A**, we obtain its row-echelon form: $\begin{pmatrix} 1 & -3 & 9 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$. The dimension of the range space of $\sigma$ is equal to the rank of **A**, which is 2. The nullity of $\sigma = 3 - 2 = 1$.

**(ii)** Since $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is in the range space of $\sigma$, there must exist $\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \in \mathbb{R}^3$ such that

$$\sigma\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \mathbf{A}\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

In other words, the following linear system in $\alpha$, $\beta$ and $\gamma$ has a solution:

$$\begin{array}{llll} 2\alpha & +\beta & +\gamma & = x \\ -\alpha & +3\beta & -9\gamma & = y \\ 3a & +\beta & +7\gamma & = z \end{array}$$

The augmented matrix $\begin{pmatrix} 2 & 1 & 1 & x \\ -1 & 3 & -9 & y \\ 3 & 1 & 7 & z \end{pmatrix}$ reduces to $\begin{pmatrix} 1 & -3 & 9 & -y \\ 0 & 1 & -2 & \frac{1}{7}(x+2y) \\ 0 & 0 & 0 & 10x-y-7z \end{pmatrix}$.

Consequently, $10x - y - 7z = 0$.

**(iii)** Since the line is the image under $\sigma$, this means that for each $\lambda$,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3+4\lambda \\ 2+5\lambda \\ 4+5\lambda \end{pmatrix}$$

is in the range. Thus $\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix}$ is in $P$ if and only if $\mathbf{A}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3+4\lambda \\ 2+5\lambda \\ 4+5\lambda \end{pmatrix}$.

From the row-echelon form obtained in **(ii)**, we have

$$\alpha - 3\beta + 9\gamma = -2 - 5\lambda$$
$$\beta - 2\gamma = 1 + 2\lambda$$

Let $\gamma = t$. Then $\beta = 1 + 2\lambda + 2t$ and $\alpha = 1 + \lambda - 3t$. Therefore, $\begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + t\begin{pmatrix} -3 \\ 2 \\ 1 \end{pmatrix}$.

This is the equation of a plane. The above vector equation can be written as $2x - y + 8z = 1$.

- Is $P$ a vector space?
- Jf $\mathrm{T}: \mathbb{R}^3 \to \mathbb{R}^3$ is a linear transformation, what is the image of a plane under T?

## §9 Eigenvalues and Eigenvectors

Many applications of matrices in both engineering and science utilize eigenvalues and, sometimes, eigenvectors. Control theory, vibration analysis, electric circuits, advanced dynamics and quantum mechanics are just a few examples of the application areas. You may refer to **Appendix III** for more details.

In Mathematics, eigenvalues and eigenvectors are used to transform a given matrix into a diagonal matrix, which helps us to evaluate powers of a square matrix.

### 9.1 Eigenvalues and Eigenvectors

**Example 9.1.1**

Consider the matrix $\mathbf{A} = \begin{pmatrix} 3 & -1 \\ -2 & 2 \end{pmatrix}$.

Find in $\mathbb{R}^2$, two nonzero and nonparallel vectors, $\mathbf{u}_1$ and $\mathbf{u}_2$, such that $\mathbf{Au}_1$ is a scalar multiple of $\mathbf{u}_1$ and $\mathbf{Au}_2$ is a scalar multiple of $\mathbf{u}_2$.

*Solution:*

Let $\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix}$. Then $\mathbf{Au} = \begin{pmatrix} 3x - y \\ -2x + 2y \end{pmatrix}$ which must be a scalar multiple of $\begin{pmatrix} x \\ y \end{pmatrix}$.

Let $\begin{pmatrix} 3x - y \\ -2x + 2y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$. Then $3x - y = \lambda x$ … (1) and $-2x + 2y = \lambda y$ … (2).

If $\lambda \neq 0$, dividing (1) by (2), we have

$$\begin{aligned} \frac{3x - y}{-2x + 2y} &= \frac{x}{y} \\ 3xy - y^2 &= -2x^2 + 2xy \\ 2x^2 + xy - y^2 &= 0 \\ (2x - y)(x + y) &= 0 \\ 2x - y = 0 &\text{ or } x + y = 0 \end{aligned}$$

If $\lambda = 0$, $3x - y = -2x + 2y = 0$ implies $x = y = 0$ (rejected as the vector is given to be nonzero)

Therefore, two such vectors can be $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ or $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Check: $\begin{pmatrix} 3 & -1 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 3 & -1 \\ -2 & 2 \end{pmatrix}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 4 \\ -4 \end{pmatrix} = 4\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

- Can you suggest some other possible answers?

> **Definition**
>
> Let **A** be an $n \times n$ matrix. A <u>nonzero</u> vector **v** in $\mathbb{R}^n$ is called an ***eigenvector*** of **A** if
>
> $$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$$
>
> for some scalar $\lambda$. The scalar $\lambda$ is called an ***eigenvalue*** of **A**, and **v** is side to be an eigenvector corresponding to $\lambda$.

In **Example 9.1.1**, $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is an eigenvector of **A** corresponding to the eigenvalue 4.

- Is it possible for a matrix to have an eigenvector **0**?
- Is it possible for a matrix to have an eigenvalue 0?

**Example 9.1.2**

Consider the matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$. Find the eigenvalues $\lambda_1$ and $\lambda_2$, and corresponding eigenvectors of **B**.

*Solution:*

Let $\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}$. Then $\mathbf{B}\mathbf{v} = \begin{pmatrix} x-2y \\ -2x+4y \end{pmatrix}$ which must be a scalar multiple of $\begin{pmatrix} x \\ y \end{pmatrix}$.

Let $\begin{pmatrix} x-2y \\ -2x+4y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$. Then $x - 2y = \lambda x \ldots$ (1) and $-2x + 4y = \lambda y \ \ldots$ (2).

If $\lambda \neq 0$, dividing (1) by (2), we have $-\frac{1}{2} = \frac{x}{y}$, i.e. $y = -2x$.

If $\lambda = 0$, $x - 2y = -2x + 4y = 0$, i.e. $x = 2y$.

A possible eigenvector is $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$, and $\begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 5 \\ -10 \end{pmatrix} = 5\begin{pmatrix} 1 \\ -2 \end{pmatrix}$, so it corresponds to the eigenvalue 5.

Another eigenvector is $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, and it corresponds to the eigenvalue 0.

This method can be very complex and tedious to apply when the size of the square matrix becomes larger.

This applet allows you to explore visually, the eigenvalues and eigenvectors of a user-defined $2 \times 2$ matrix: https://www.geogebra.org/m/KuMAuEnd.

The following theorem can help us simplify the process of finding the eigenvalue(s) of a square matrix.

**Theorem 9.1.1**

Let $\mathbf{A}$ be an $n \times n$ matrix. Then $\lambda$ is an eigenvalue of $\mathbf{A}$ if and only if $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$.

*Proof:*

$\lambda$ is an eigenvalue of $\mathbf{A}$.

$\Leftrightarrow$ There exists a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ (the eigenvector) such that $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$.

$\Leftrightarrow$ There exists a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ (the eigenvector) such that $\lambda\mathbf{v} - \mathbf{A}\mathbf{v} = \mathbf{0}$.

$\Leftrightarrow$ There exists a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ (the eigenvector) such that $\lambda\mathbf{I}\mathbf{v} - \mathbf{A}\mathbf{v} = \mathbf{0}$.

$\Leftrightarrow$ There exists a nonzero vector $\mathbf{v} \in \mathbb{R}^n$ (the eigenvector) such that $(\lambda\mathbf{I} - \mathbf{A})\mathbf{v} = \mathbf{0}$.

$\Leftrightarrow$ The homogeneous linear system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{v} = \mathbf{0}$ has a nontrivial solution.

$\Leftrightarrow$ $\det(\lambda\mathbf{I} - \mathbf{A}) = 0$.

To find an eigenvector corresponding to a found eigenvalue, is equivalent to find a nontrivial solution of the homogeneous linear system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{v} = \mathbf{0}$.

**Example 9.1.3**

Use **Theorem 9.1.1** to find all the eigenvalues and the corresponding eigenvectors of $\begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$.

*Solution:*

$$\lambda\mathbf{I} - \mathbf{B} = \begin{pmatrix} \lambda - 1 & 2 \\ 2 & \lambda - 4 \end{pmatrix}$$

$$0 = \det(\lambda\mathbf{I} - \mathbf{B}) = \begin{vmatrix} \lambda - 1 & 2 \\ 2 & \lambda - 4 \end{vmatrix} = (\lambda - 1)(\lambda - 4) - 4 = \lambda^2 - 5\lambda$$

So $\lambda = 0$ or $\lambda = 5$.

When $\lambda = 0$, we solve $\begin{pmatrix} -1 & 2 \\ 2 & -4 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ to find eigenvector. By observation, a corresponding eigenvector $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$.

When $\lambda = 5$, we solve $\begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ to find eigenvector. By observation, a corresponding eigenvector $\begin{pmatrix} -1 \\ 2 \end{pmatrix}$.

**Definition**

Let **A** be an $n \times n$ matrix. The equation

$$\det(\lambda\mathbf{I} - \mathbf{A}) = 0$$

is called the ***characteristic equation*** of **A**. $\det(\lambda\mathbf{I} - \mathbf{A})$, when expanded, is a polynomial in $\lambda$, and is called the ***characteristic polynomial*** of **A**.

- What can you say about the number of eigenvalues that a square matrix has?

**Example 9.1.4**

Let $\mathbf{A} = \begin{pmatrix} 1 & 2 & -1 \\ 1 & 0 & 1 \\ 4 & -4 & 5 \end{pmatrix}$.

**(a)** Find all the eigenvalues of **A**.
**(b)** Find an eigenvector corresponding to each eigenvalue in **(a)**.

*Solution:*

**(a)**

$\det(\lambda\mathbf{I} - \mathbf{A}) = 0$ gives $\begin{vmatrix} \lambda - 1 & -2 & 1 \\ -1 & \lambda & -1 \\ -4 & 4 & \lambda - 5 \end{vmatrix} = 0$

$$(\lambda-1)(\lambda)(\lambda-5)+(-1)(4)(1)+(-4)(-2)(-1)-(-4)(\lambda)(1)-(-1)(-2)(\lambda-5)-(\lambda-1)(4)(-1)=0$$

$$\lambda^3 - 6\lambda^2 + 5\lambda - 4 - 8 + 4\lambda - 2\lambda + 10 + 4\lambda - 4 = 0$$

$$\lambda^3 - 6\lambda^2 + 11\lambda - 6 = 0$$

$$(\lambda-1)(\lambda-2)(\lambda-3) = 0$$

Therefore the eigenvalues of **A** are 1, 2 and 3.

**(b)**

When $\lambda = 1$, a nontrivial solution of $\begin{pmatrix} 0 & -2 & 1 \\ -1 & 1 & -1 \\ -4 & 4 & -4 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 1 \\ 2 \end{pmatrix}$ is a corresponding eigenvector.

When $\lambda = 2$, a nontrivial solution of $\begin{pmatrix} 1 & -2 & 1 \\ -1 & 2 & -1 \\ -4 & 4 & -3 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -2 \\ 1 \\ 4 \end{pmatrix}$ is a corresponding eigenvector.

When $\lambda = 3$, a nontrivial solution of $\begin{pmatrix} 2 & -2 & 1 \\ -1 & 3 & -1 \\ -4 & 4 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 1 \\ 4 \end{pmatrix}$ is a corresponding eigenvector.

- What can you say about the eigenvalues and eigenvectors of the matrix $\mathbf{A} + 2\mathbf{I}$ ?

**Eigenspace** *(not in H2 FM syllabus)*

**Theorem 9.1.2**

Let $\mathbf{A}$ be an $n \times n$ matrix and let $\lambda$ be an eigenvalue of $\mathbf{A}$. Let $E_\lambda$ denote the set of all eigenvectors of $\mathbf{A}$ corresponding to the eigenvalue $\lambda$, together with the zero vector $\mathbf{0}$. In other words,

$$E_\lambda = \left\{ \mathbf{v} \in \mathbb{R}^n : \mathbf{Av} = \lambda \mathbf{v} \right\}.$$

Then $E_\lambda$ is the null space of $(\lambda \mathbf{I} - \mathbf{A})$.

*Proof:*

Note that for $\mathbf{v} \in \mathbb{R}^n$,

$$\mathbf{v} \in E_\lambda \Leftrightarrow \mathbf{Av} = \lambda \mathbf{v} \Leftrightarrow (\lambda \mathbf{I} - \mathbf{A})\mathbf{v} = \mathbf{0} \Leftrightarrow \mathbf{v} \in \text{null space of } (\lambda \mathbf{I} - \mathbf{A}).$$

Therefore, $E_\lambda = \text{null space of } (\lambda \mathbf{I} - \mathbf{A})$.

**Definition**

Consequently, $E_\lambda = \left\{ \mathbf{v} \in \mathbb{R}^n : \mathbf{Av} = \lambda \mathbf{v} \right\}$ as a subspace of $\mathbb{R}^n$ is called the ***eigenspace*** of $\mathbf{A}$ corresponding to the eigenvalue $\lambda$.

**Example 9.1.5**

Determine whether the following statement is true:

"Let $\mathbf{A}$ be an $n \times n$ matrix and let $\lambda$ be an eigenvalue of $\mathbf{A}$. If $\mathbf{u}$ and $\mathbf{v}$ are two eigenvectors corresponding to $\lambda$, then they must be parallel, i.e. one is a scalar multiple of another."

Justify your answer.

*Solution:*

This is only true if the dimension of the eigenspace of $\mathbf{A}$ corresponding to $\lambda$ is 1, so the statement is false.

Let us construct a counterexample: Suppose we want to construct a $3 \times 3$ matrix $(\lambda \mathbf{I} - \mathbf{A})$, such that its null space has dimension 2, this implies that its rank must be 1.

One such matrix can be $\begin{pmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \\ 3 & 0 & 6 \end{pmatrix}$. Two possible nonparallel eigenvectors can be $\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$.

To check:

We let $\lambda = 1$. Then $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \\ 3 & 0 & 6 \end{pmatrix} = \begin{pmatrix} 0 & 0 & -2 \\ -2 & 1 & -4 \\ -3 & 0 & -5 \end{pmatrix}$.

$\begin{pmatrix} 0 & 0 & -2 \\ -2 & 1 & -4 \\ -3 & 0 & -5 \end{pmatrix}\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} = 1\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 0 & -2 \\ -2 & 1 & -4 \\ -3 & 0 & -5 \end{pmatrix}\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = 1\begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$.

## 9.2 Diagonalization

In many applications, it is desired to find the *n*th power of a given matrix **A**. If **A** is a diagonal matrix, then it is relatively easy to compute $\mathbf{A}^n$.

**Theorem 9.2.1**

Let **A** be an $m \times m$ diagonal matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{mm} \end{pmatrix}.$$

Then its *m*th power

$$\mathbf{A}^n = \begin{pmatrix} a_{11}{}^n & 0 & \cdots & 0 \\ 0 & a_{22}{}^n & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{mm}{}^n \end{pmatrix}.$$

*Proof:*

The result can be proven by mathematical induction (omitted).

- What if **A** is not diagonal?

**Example 9.2.1**

From **Example 9.1.1**, it is known that $\mathbf{A} = \begin{pmatrix} 3 & -1 \\ -2 & 2 \end{pmatrix}$ has eigenvectors $\mathbf{u}_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\mathbf{u}_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ corresponding to the eigenvalues 1 and 4 respectively. Let

$$\mathbf{P} = \left(\mathbf{u}_1 \mid \mathbf{u}_2\right) = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} \text{ and } \mathbf{D} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}.$$

**(i)** Verify that $\mathbf{AP} = \mathbf{PD}$.
**(ii)** Prove that **(i)** is true for any general $2 \times 2$ matrix.

*Proof:*

**(i)** $\mathbf{AP}=\begin{pmatrix}3 & -1\\ -2 & 2\end{pmatrix}\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}=\begin{pmatrix}1 & 4\\ 2 & -4\end{pmatrix}$, and $\mathbf{PD}=\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 4\end{pmatrix}=\begin{pmatrix}1 & 4\\ 2 & -4\end{pmatrix}$. Thus $\mathbf{AP}=\mathbf{PD}$.

**(ii)** $\mathbf{AP}=\mathbf{A}\left(\mathbf{u}_1 \mid \mathbf{u}_2\right)=\left(\mathbf{Au}_1 \mid \mathbf{Au}_2\right)$, and $\mathbf{PD}=\left(\mathbf{u}_1 \mid \mathbf{u}_2\right)\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}=\left(\lambda_1\mathbf{u}_1 \mid \lambda_2\mathbf{u}_2\right)$. Since $\mathbf{u}_1$ and $\mathbf{u}_1$ are eigenvectors of **A** corresponding to the eigenvalues $\lambda_1$ and $\lambda_2$ respectively,

$$\mathbf{Au}_1=\lambda_1\mathbf{u}_1 \text{ and } \mathbf{Au}_2=\lambda_2\mathbf{u}_2.$$

Thus $\mathbf{AP}=\mathbf{PD}$.

- Can you extend the proof for a general $m\times m$ matrix?

Note that $\mathbf{AP}=\mathbf{PD}\Rightarrow\mathbf{A}=\mathbf{PDP}^{-1}$ if **P** is invertible. In this case

$$\mathbf{A}^m=\left(\mathbf{PDP}^{-1}\right)^m=\underbrace{\left(\mathbf{PDP}^{-1}\right)\left(\mathbf{PDP}^{-1}\right)\ldots\left(\mathbf{PDP}^{-1}\right)}_{m\text{ times}}=\mathbf{PD}^m\mathbf{P}^{-1}.$$

**Example 9.2.2**

Use the above result to find $\mathbf{A}^5$ where $\mathbf{A}=\begin{pmatrix}3 & -1\\ -2 & 2\end{pmatrix}$.

*Solution:*

$\mathbf{P}=\left(\mathbf{u}_1 \mid \mathbf{u}_2\right)=\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}$ and $|\mathbf{P}|=-3$, so $\mathbf{P}^{-1}=-\frac{1}{3}\begin{pmatrix}-1 & -1\\ -2 & 1\end{pmatrix}=\frac{1}{3}\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}$. $\mathbf{D}=\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 4\end{pmatrix}$.

Now,

$$\begin{aligned}\mathbf{A}^5&=\mathbf{PD}^5\mathbf{P}^{-1}\\ &=\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}\begin{pmatrix}1^5 & 0\\ 0 & 4^5\end{pmatrix}\left[\frac{1}{3}\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}\right]\\ &=\frac{1}{3}\begin{pmatrix}1 & 1024\\ 2 & -1024\end{pmatrix}\begin{pmatrix}1 & 1\\ 2 & -1\end{pmatrix}\\ &=\frac{1}{3}\begin{pmatrix}2049 & -1023\\ -2046 & 1026\end{pmatrix}\\ &=\begin{pmatrix}683 & -341\\ -682 & 342\end{pmatrix}\end{aligned}.$$

- How do you check the answer effectively?

**Definition**

A square matrix $\mathbf{A}$ is called ***diagonalizable*** if there is an invertible matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is a diagonal matrix. The matrix $\mathbf{P}$ is said to diagonalize $\mathbf{A}$.

Note that the order of matrix multiplication is important in the results:

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \text{ and } \mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}.$$

**Theorem 9.2.2**

If $\mathbf{A}$ is an $n \times n$ matrix, then $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ has $n$ linearly independent eigenvectors.

*Proof:*

Suppose $\mathbf{A}$ is is diagonalizable. Then by definition there exists an invertible matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix, so $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$.

Let $\mathbf{P} = (\mathbf{v}_1 \mid \mathbf{v}_2 \mid \ldots \mid \mathbf{v}_n)$ and $\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$. Then $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$ implies

$$(\mathbf{A}\mathbf{v}_1 \mid \mathbf{A}\mathbf{v}_2 \mid \ldots \mid \mathbf{A}\mathbf{v}_n) = (d_1\mathbf{v}_1 \mid d_2\mathbf{v}_2 \mid \ldots \mid d_n\mathbf{v}_n), \text{ i.e.}$$
$$\mathbf{A}\mathbf{v}_1 = d_1\mathbf{v}_1,\ \mathbf{A}\mathbf{v}_2 = d_2\mathbf{v}_2, \ldots, \mathbf{A}\mathbf{v}_n = d_n\mathbf{v}_n.$$

These vectors are eigenvectors of $\mathbf{A}$ by definition.

Since $\mathbf{P}$ is invertible, its nullity is 0. Thus its rank (or the dimension of its column space) is $n$. Therefore $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ must be a basis of its column space thus these vectors are linearly independent.

Suppose $\mathbf{A}$ has $n$ linearly independent vectors, say $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively. Let $\mathbf{P}$ be the matrix $(\mathbf{v}_1 \mid \mathbf{v}_2 \mid \ldots \mid \mathbf{v}_n)$ and $\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$.

Since the $n$ eigenvectors are linearly independent, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of the column space of $\mathbf{P}$. The matrix $\mathbf{P}$ is invertible as its rank is $n$ and its nullity is 0. Also we have

$$\mathbf{A}\mathbf{P} = (\mathbf{A}\mathbf{v}_1 \mid \mathbf{A}\mathbf{v}_2 \mid \ldots \mid \mathbf{A}\mathbf{v}_n) = (\lambda_1\mathbf{v}_1 \mid \lambda_2\mathbf{v}_2 \mid \ldots \mid \lambda_n\mathbf{v}_n) = \mathbf{P}\mathbf{D},$$

so $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$. Hence $\mathbf{A}$ is diagonalizable by definition.

**Theorem 9.2.3**

If an $n \times n$ matrix **A** has $n$ distinct eigenvalues, then A is diagonalizable.

**Important:**

**Theorem 9.2.3** gives a sufficient condition but not a necessary condition for **A** to be diagonalizable.

- Can you give an example, in which an $n \times n$ matrix **A** does not have $n$ distinct eigenvalues but it is still diagonalizable?

**Example 9.2.3**

For a $3 \times 3$ matrix **B** whose eigenvalues are 1, $-2$ and $-3$, and for which corresponding eigenvectors are $\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ respectively,

The element in the first row and the first column of $\mathbf{B}^n$ is denoted by $\alpha$. Show that

$$\alpha = \frac{(-2)^n + (-3)^n}{2}.$$

*Proof:*

Let $\mathbf{B} = \mathbf{PDP}^{-1}$, where $\mathbf{P} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ and $\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{pmatrix}$. $\mathbf{P}^{-1} = \frac{1}{2}\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$

Thus $\mathbf{B}^n = \mathbf{PD}^n\mathbf{P}^{-1} = \frac{1}{2}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & (-2)^n & 0 \\ 0 & 0 & (-3)^n \end{pmatrix}\begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$

$\alpha = (\text{first row of } \mathbf{PD})(\text{first column of } \mathbf{P}^{-1}) = (\text{first row of } \mathbf{P})\mathbf{D}(\text{first column of } \mathbf{P}^{-1})$

$$= \frac{1}{2}\begin{pmatrix} 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & (-2)^n & 0 \\ 0 & 0 & (-3)^n \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 0 & (-2)^n & (-3)^n \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} = \frac{(-2)^n + (-3)^n}{2}$$

**Example 9.2.4**

Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \\ 0 & 1 & 3 \end{pmatrix}$. Determine whether **A** is diagonalizable, and find an invertible matrix **P** and a diagonal matrix **D** such that $\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D}$ if so.

*Solution:*

The characteristic equation of **A** is

$$\det(\lambda\mathbf{I}-\mathbf{A})=\begin{vmatrix}\lambda-1 & -1 & -2\\ 0 & \lambda-1 & 0\\ 0 & -1 & \lambda-3\end{vmatrix}=(\lambda-1)^2(\lambda-3)=0$$, the eigenvalues of **A** are 1 and 3.

When $\lambda=1$, $\mathbf{I}-\mathbf{A}=\begin{pmatrix}0 & -1 & -2\\ 0 & 0 & 0\\ 0 & -1 & -2\end{pmatrix}\to\begin{pmatrix}0 & 1 & 2\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}$. To find the nontrivial solutions of

$$\begin{pmatrix}0 & 1 & 2\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$, i.e. $y+2z=0$,

we let $x=s$ and $z=t$, so $y=-2t$. Then $\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}s\\ -2t\\ t\end{pmatrix}=s\begin{pmatrix}1\\ 0\\ 0\end{pmatrix}+t\begin{pmatrix}0\\ -2\\ 1\end{pmatrix}$. Thus, two independent eigenvectors, corresponding to the eigenvalue 1, are $\begin{pmatrix}1\\ 0\\ 0\end{pmatrix}$ and $\begin{pmatrix}0\\ -2\\ 1\end{pmatrix}$.

When $\lambda=3$, $3\mathbf{I}-\mathbf{A}=\begin{pmatrix}2 & -1 & -2\\ 0 & 2 & 0\\ 0 & -1 & 0\end{pmatrix}\to\begin{pmatrix}1 & 0 & -1\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix}$. To find the nontrivial solutions of

$$\begin{pmatrix}1 & 0 & -1\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$, i.e. $x-z=0$ and $y=0$,

we let $z=r$, so $x=-r$. Then $\begin{pmatrix}x\\ y\\ z\end{pmatrix}=\begin{pmatrix}r\\ 0\\ r\end{pmatrix}=r\begin{pmatrix}1\\ 0\\ 1\end{pmatrix}$. Thus, an eigenvector corresponding to the eigenvalue 3, is $\begin{pmatrix}1\\ 0\\ 1\end{pmatrix}$.

Now we construct $\mathbf{P}=\begin{pmatrix}1 & 0 & 1\\ 0 & -2 & 0\\ 0 & 1 & 1\end{pmatrix}$. Since it is invertible, $\mathbf{D}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 3\end{pmatrix}=\mathbf{P}^{-1}\mathbf{AP}$.

- In this example, what can you say about the sum of the dimensions of all the eigenspaces of **A**?

Here is another necessary and sufficient condition for a square matrix to be diagonalizable. (*not in H2 FM Syllabus*)

**Theorem 9.2.4**

Let **A** be an $n \times n$ matrix. Then **A** is diagonalizable if and only if the sum of the dimensions of all the eigenspaces of **A** is *n*. That is, if $\lambda_1, \lambda_2, ..., \lambda_k$ ($k \le n$) are the distinct eigenvalues of **A**, then **A** is diagonalizable if and only if

$$\dim\left(E_{\lambda_1}\right) + \dim\left(E_{\lambda_2}\right) + ... + \dim\left(E_{\lambda_k}\right) = n .$$

In Example **9.2.4**, when $\lambda = 1$, $\text{rank}\left(\mathbf{I} - \mathbf{A}\right) = 1$ so $\dim\left(E_1\right) = \text{nullity}\left(\mathbf{I} - \mathbf{A}\right) = 3 - 1 = 2$; when $\lambda = 3$, $\text{rank}\left(3\mathbf{I} - \mathbf{A}\right) = 2$, so $\dim\left(E_3\right) = \text{nullity}\left(3\mathbf{I} - \mathbf{A}\right) = 3 - 2 = 1$.

Since $\dim\left(E_1\right) + \dim\left(E_3\right) = 3 = n$, **A** is diagonalizable.

## 9.3 Application to Linear Recurrence Relations

We illustrate with an example the application of diagonalization to solving some linear recurrence relations.

**Example 9.3.1**

A sequence of numbers $a_0$, $a_1$, $a_2$, … is defined by the linear recurrence relation

$$a_n = a_{n-1} + 6a_{n-2}, \; n \ge 2 .$$

Let the column vector $\mathbf{u}_n$ denote $\begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix}$.

**(i)** Find a $2 \times 2$ matrix **A** such that $\mathbf{u}_n = \mathbf{A}\mathbf{u}_{n-1}$.

**(ii)** Hence, express $\mathbf{u}_n$ in the form $\mathbf{B}\mathbf{u}_0$, where **B** is a $2 \times 2$ matrix to be determined.

**(iii)** Deduce the expression of $a_n$ in terms of $a_1$, $a_0$ and *n*.

*Solution:*

**(i)** $\mathbf{u}_n = \begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} = \begin{pmatrix} a_n + 6a_{n-1} \\ a_n + 0a_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 6 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a_n \\ a_{n-1} \end{pmatrix}$, so $\mathbf{A} = \begin{pmatrix} 1 & 6 \\ 1 & 0 \end{pmatrix}$.

**(ii)** $\mathbf{u}_n = \mathbf{A}\mathbf{u}_{n-1} = \mathbf{A}\left(\mathbf{A}\mathbf{u}_{n-2}\right) = ... = \left(\mathbf{A}^n\right)\mathbf{u}_0$, so $\mathbf{B} = \mathbf{A}^n$. Now we need to diagonalize **A**:

$\det\left(\lambda\mathbf{I} - \mathbf{A}\right) = \begin{vmatrix} \lambda - 1 & -6 \\ -1 & \lambda \end{vmatrix} = \lambda^2 - \lambda - 6 = 0$, so the eigenvalues are $-2$ and 3.

When $\lambda = 2$, we solve $\begin{pmatrix} -3 & -6 \\ -1 & -2 \end{pmatrix}\mathbf{v} = \mathbf{0}$, a corresponding eigenvector is $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$.

When $\lambda = 3$, we solve $\begin{pmatrix} 2 & -6 \\ -1 & 3 \end{pmatrix}\mathbf{v} = \mathbf{0}$, a corresponding eigenvector is $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$.

Now we can write $\mathbf{A} = \mathbf{PDP}^{-1}$, where $\mathbf{P} = \begin{pmatrix} 2 & 3 \\ -1 & 1 \end{pmatrix}$ and $\mathbf{D} = \begin{pmatrix} -2 & 0 \\ 0 & 3 \end{pmatrix}$. $\mathbf{P}^{-1} = \frac{1}{5}\begin{pmatrix} 1 & -3 \\ 1 & 2 \end{pmatrix}$.

$$\begin{aligned}
\mathbf{A}^n &= \mathbf{PD}^n\mathbf{P}^{-1} \\
&= \frac{1}{5}\begin{pmatrix} 2 & 3 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} (-2)^n & 0 \\ 0 & 3^n \end{pmatrix}\begin{pmatrix} 1 & -3 \\ 1 & 2 \end{pmatrix} \\
&= \frac{1}{5}\begin{pmatrix} 2(-2)^n & 3(3^n) \\ -1(-2)^n & (3^n) \end{pmatrix}\begin{pmatrix} 1 & -3 \\ 1 & 2 \end{pmatrix} \\
&= \frac{1}{5}\begin{pmatrix} 2(-2)^n + 3(3^n) & -6(-2)^n + 6(3^n) \\ -(-2)^n + (3^n) & 3(-2)^n + 2(3^n) \end{pmatrix}
\end{aligned}$$

So $\mathbf{B} = \frac{1}{5}\begin{pmatrix} 2(-2)^n + 3(3^n) & -6(-2)^n + 6(3^n) \\ -(-2)^n + (3^n) & 3(-2)^n + 2(3^n) \end{pmatrix}$.

**(iii)** $$\begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 2(-2)^n + 3(3^n) & -6(-2)^n + 6(3^n) \\ -(-2)^n + (3^n) & 3(-2)^n + 2(3^n) \end{pmatrix}\begin{pmatrix} a_1 \\ a_0 \end{pmatrix}$$

$$= \frac{1}{5}\begin{pmatrix} \left[2(-2)^n + 3(3^n)\right]a_1 + \left[-6(-2)^n + 6(3^n)\right]a_0 \\ \left[-(-2)^n + (3^n)\right]a_1 + \left[3(-2)^n + 2(3^n)\right]a_0 \end{pmatrix}$$

$$= \frac{1}{5}\begin{pmatrix} (2a_1 - 6a_0)(-2)^n + (3a_1 + 6a_0)(3^n) \\ (-a_1 + 3a_0)(-2)^n + (a_1 + 2a_0)(3^n) \end{pmatrix}$$

Therefore $a_n = \dfrac{(-a_1 + 3a_0)(-2)^n + (a_1 + 2a_0)(3^n)}{5}$.

This approach can be extended to solve a higher-order linear homogeneous recurrence relation, and even differential equations. Refer to **Appendix III** for more details.

# SUMMARY PAGE 1

# SUMMARY PAGE 2

## SUMMARY PAGE 3

## Appendix I: Calculators

### 1.1 Commands of Graphic Calculator (TI-84c)

| Menu | Details |
|---|---|
| 1: det( | 'det([A])' returns the determinant of square matrix [A]. |
| 2: $^{T}$ | $[A]^T$ returns the transpose of matrix [A]. |
| 3: dim ( | 'dim([A])' returns the size of matrix [A] |
| 4: Fill ( | 'Fill($a$,[A])' fills / replaces all the elements of [A] with $a$. |
| 5: identity( | 'identity($n$)' returns a $n \times n$ identity matrix. |
| 6: randM( | 'randM($m$,$n$)' returns a random $m \times n$ matrix (integer elements from -9 to 9). |
| 7: augment( | 'augment([A],[B])' appends matrices [A] and [B] together. |
| 8: Matr>list( | 'Matr>list([A],$L_1$,$L_2$,…)' fills each of the list with the columns of [A], neglecting excess. |
| 9: List>matr( | 'List>matr($L_1$,$L_2$,…,[A])' fills each column of [A] with the lists, neglecting excess. |
| 0: cumSum( | 'cumSum([A])' returns the cumulative sums of a matrix. |
| A: ref( | 'ref([A])' returns a row-echelon form of matrix [A]. |
| B: rref( | 'rref([A])' returns the reduced row-echelon form of matrix [A]. |
| C: rowSwap( | 'rowSwap([A],$i$,$j$)' returns the matrix obtained by swapping rows $i$ and $j$ in [A]. |
| D: row+( | 'row+([A],$i$,$j$)' returns the matrix obtained by adding row $i$ to row $j$ in [A]. |
| E: *row( | *row($k$,[A],$i$)' returns the matrix obtained by multiplying row $i$ in [A] by $k$. |
| F: *row+( | '*row+($k$,[A],$i$,$j$)' returns the matrix obtained by adding $k$ times row $i$ to row $j$ in [A]. |

*The highlighted commands are not required in H2 FM Syllabus.*

### 1.2 Online Calculators

**(a)** An online calculator for matrices
URL: http://matrix.reshish.com/

**(b)** An online calculator for eigenvalues and eigenvectors:
URL: http://www.mathportal.org/calculators/matrices-calculators/matrix-calculator.php

**(c)** Explore and record other online calculators yourself:

## Appendix II: Some Mathematical Terminologies

*Definition* - a precise and unambiguous description of the meaning of a mathematical term. It characterizes the meaning of a word by giving all the properties and only those properties that must be true.
e.g. definition of elementary row operations.

*Theorem* - a mathematical statement that is proved using rigorous mathematical reasoning. In a mathematical paper, the term theorem is often reserved for the most important results.
e.g. Pythagoras Theorem.

*Lemma* - a minor result whose sole purpose is to help in proving a theorem. It is a stepping stone on the path to proving a theorem. Very occasionally lemmas can take on a life of their own.
e.g. Zorn's lemma, Urysohn's lemma, Burnside's lemma, Sperner's lemma.

*Corollary* - a result in which the (usually short) proof relies heavily on a given theorem. We often say that "this is a corollary of Theorem A".
e.g. the corollaries in **Section 4**.

*Proposition* - a proven and often interesting result, but generally less important than a theorem.
e.g. some statements that you have shown by mathematical induction.

*Conjecture* - a statement that is unproved, but is believed to be true.
e.g. Collatz conjecture, Goldbach conjecture, twin prime conjecture.

*Axiom/*Postulate - a statement that is assumed to be true without proof. These are the basic building blocks from which all theorems are proven.
e.g. Euclid's five postulates, Zermelo-Fraenkel axioms, Peano axioms.

*Identity -* a mathematical expression giving the equality of two (often variable) quantities.
e.g. trigonometric identities, Euler's identity.

*Paradox* - a statement that can be shown, using a given set of axioms and definitions, to be both true and false. Paradoxes are often used to show the inconsistencies in a flawed theory (Russell's paradox). The term paradox is often used informally to describe a surprising or counterintuitive result that follows from a given set of rules.
e.g. Banach-Tarski paradox, Alabama paradox, Gabriel's horn.

## Appendix III: Some Applications

The following online resources are selected from *Linear Algebra Larson 7th Edition*.

**3.1 System of Linear Equations** (url: http://tinyurl.com/MandLapp1)**:**

**(a)** Set up and solve a system of equations to fit a polynomial function to a set of data points.
**(b)** Set up and solve a system of equations to represent a network.

**3.2 Applications of Matrix Operations** (url: http://tinyurl.com/MandLapp2)**:**

**(a)** Write and use a stochastic matrix.
**(b)** Use matrix multiplication to encode and decode messages.
**(c)** Use matrix algebra to analyse an economic system (Leontief input-output model).
**(d)** Find the least squares regression line for a set of data.

**3.3 Applications of Determinants** (url: http://tinyurl.com/MandLapp3)**:**

**(a)** Find the adjoint of a matrix and use it to find the inverse of the matrix.
**(b)** Use Cramer's Rule to solve a system of $n$ linear equations in $n$ variables.
**(c)** Use determinants to find area, volume, and the equations of lines and planes.

**3.4 Applications of Vector Spaces** (url: http://tinyurl.com/MandLapp4)**:**

**(a)** Use the Wronskian to test a set of solutions of a linear homogeneous differential equation for linear independence.
**(b)** Identify and sketch the graph of a conic section and perform a rotation of axes.

**3.5 Applications of Inner Product Spaces** (url: http://tinyurl.com/MandLapp5)**:**

**(a)** Find the cross product of two vectors in $\mathbb{R}^3$.
**(b)** Find the linear or quadratic least square approximation of a function.
**(c)** Find the $n$th-order Fourier approximation of a function.

**3.6 Applications of Linear Transformations** (url: http://tinyurl.com/MandLapp6)**:**

**(a)** Identify linear transformations defined by reflections, expansions, contracts, or shears in $\mathbb{R}^2$.
**(b)** Use a linear transformation to rotate a figure in $\mathbb{R}^3$.

**3.7 Applications of Eigenvalues and Eigenvectors** (url: http://tinyurl.com/MandLapp7)**:**

**(a)** Model population growth using an age transition matrix and an age distribution vector, and find a stable age distribution vector.
**(b)** Use a matrix equation to solve a system of first-order linear differential equations.
**(c)** Find the matrix of a quadratic form and use the Principal Axes Theorem to perform a rotation of axes for a conic and a quadric surface.

**3.8 Record any resources that you have found out:**